

Durham Research Online

Deposited in DRO:

22 August 2017

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Karagiannis, G. and Konomi, B. and Lin, G. (2015) 'A Bayesian mixed shrinkage prior procedure for spatial-stochastic basis selection and evaluation of gPC expansions : applications to elliptic SPDEs.', *Journal of computational physics.*, 284 . pp. 528-546.

Further information on publisher's website:

<https://doi.org/10.1016/j.jcp.2014.12.034>

Publisher's copyright statement:

© 2015 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

A Bayesian mixed shrinkage prior procedure for spatial-stochastic basis selection and evaluation of gPC expansions: Applications to elliptic SPDEs

Georgios Karagiannis^a, Bledar A. Konomi^b, Guang Lin^{a,*}

^a*Department of Mathematics, School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA*

^b*Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221, USA*

Abstract

We propose a new fully Bayesian method to efficiently obtain the spectral representation of a spatial random field, which can conduct spatial-stochastic basis selection and evaluation of generalized Polynomial Chaos (gPC) expansions when the number of the available basis functions is significantly larger than the size of the training data-set. We develop a fully Bayesian stochastic procedure, called mixed shrinkage prior (MSP), which performs both basis selection and coefficient evaluation simultaneously. MSP involves assigning a prior probability on the gPC structure and assigning conjugate priors on the expansion coefficients that can be thought of as mixtures of Ridge-LASSO shrinkage priors, in augmented form. The method offers a number of advantages over existing compressive sensing methods in gPC literature, such that it recovers possible sparse structures in both stochastic and spatial domains while the resulted expansion can be re-used directly to economically obtain results at any spatial input values. Yet, it inherits all the advantages of Bayesian model uncertainty methods, e.g. accounts for uncertainty about basis significance and provides interval estimation through posterior distributions. A unique highlight of the MSP procedure is that it can address heterogeneous sparsity in the spatial domain for different random dimensions. Furthermore, it yields a compromise between Ridge and LASSO regressions, and hence combines a weak (l_2 -norm) and strong (l_1 -norm) shrinkage, in an adaptive, data-driven manner. We demonstrate the good performance of the proposed method, and compare it against other existing compressive sensing ones on elliptic stochastic partial differential equations.

Keywords: Generalized polynomial chaos, uncertainty quantification, stochastic PDE, sparse approximation, Bayesian model uncertainty, MCMC

*Corresponding author

Email addresses: gkaragia@purdue.edu (Georgios Karagiannis), alex.konomi@uc.edu (Bledar A. Konomi), guanglin@purdue.edu (Guang Lin)

1. Introduction

Uncertainty quantification (UQ) provides a quantitative characterization of uncertainties in stochastic systems and the efficient propagation of them for model prediction given available data. In physical systems, for example, the mathematical models may be governed by stochastic partial differential equations (SPDEs) that include random and spatial input variables. A mathematical representation of the stochastic solution with respect to the input variables is usually of interest. In many cases, such a mapping does not have an explicit analytical form. This requires the construction and use of a surrogate model, which is a tractable but approximated mathematical expression of this mapping.

Monte Carlo (MC) methods (Metropolis et al., 1953; Ripley, 1987) have been extensively used in UQ. MC methods allow for the evaluation of solution statistics from a sample of solutions collected after several system realizations. Each system realization is performed by drawing a random input value from a given distribution and running a solver for the system. Because they require many system realizations (MC samples) (Xiu, 2010), MC methods are generally impractical for large-scale systems if the solver is computationally expensive to run.

The generalized Polynomial Chaos (gPC) methods (Xiu, 2010; Xiu and Karniadakis, 2002; Ghanem and Spanos, 2003; Wan and Karniadakis, 2005, 2006) have been successively applied to a variety of UQ problems as they can provide satisfactory surrogate models. Briefly, in the gPC context, the stochastic solution, taking as arguments random and spatial input variables, is expanded as a convergent series of PC bases and PC coefficients which include the random and spatial input dimensions correspondingly. The toolbox of traditional computational approaches includes the stochastic Galerkin projection methods (Ghanem and Spanos, 2003; Deb et al., 2001; Babuška et al., 2004) and stochastic collocation methods (Mathelin and Hussaini, 2003; Babuška et al., 2007; Xiu and Hesthaven, 2005). Although these computational approaches require less system evaluations than MC, they suffer from the curse of dimensionality.

Curse of dimensionality: When the dimension of the random input variables increases, the number of unknown PC coefficients increases dramatically due to the tensor product involved in the design of the multivariate PC bases. In practice, only a small number of training data may be available, perhaps even smaller than the number of the unknown PC coefficients. Hence, the evaluation problem may become under-determined, causing traditional computational approaches to give unstable and inaccurate results due to over-fitting (Doostan and Owhadi, 2011; Yang and Karniadakis, 2012). Furthermore, in this context, the consideration of the PC coefficients as functions over the whole spatial input –not just at a fixed spatial input value– makes the evaluation of the gPC expansion even more challenging.

In elliptic PDEs with high-dimensional random diffusion coefficients, the stochastic solution can be sparse in the stochastic domain (Babuška et al., 2007; Doostan and Owhadi, 2011; Yang and Karniadakis, 2012; Todor and Schwab, 2007; Bieri and Schwab, 2009; Bieri et al., 2009) when expanded as a gPC; namely only a very small fraction of available PC bases are dominant and able to recover the stochastic solution without any significant loss in accuracy. We will refer to these PC bases as significant. Here, the curse of dimensionality occurs as a problem because the deterministic solver is expensive to run. Thus, the number of available training data is limited. One remedy is to reduce the number of the PC bases –and hence, the number of the unknown PC coefficients– by omitting those whose absence does not cause unreliable results (Doostan and Owhadi, 2011).

To address the curse of dimensionality issue, dimension reduction techniques, such as compressive sensing, have been proposed in the literature (Doostan and Owhadi, 2011; Yang and Karniadakis, 2012). These techniques aim to approximate the solution economically, given a small MC sample and without sacrificing accuracy, by shrinking the values of the PC coefficients towards zero in order to retain only the significant PC bases. In particular, Doostan and Owhadi (2011) suggested the use of compressive sensing techniques based on l_1 -minimization algorithm, while Yang and Karniadakis (2012) improved this idea computationally by employing a re-weighted l_1 -minimization algorithm. These methods are non-adaptive, provably convergent and well suited for high-dimensional cases with a limited number of MC samples. The use of the Bayesian compressive sensing algorithm (Ji et al., 2008, 2009) as an alternative approach was suggested by Sargsyan et al. (2013). These methods strictly require the solution to be sparse. Other approaches include, ANOVA decomposition (Yang et al., 2012; Ma and Zabaras, 2010; Li et al., 2014; Wei et al., in press), sparse decomposition for the stochastic Galerkin technique (Bieri et al., 2009), and adaptive sparse grids for the stochastic collocation (Ma and Zabaras, 2009; Nobile et al., 2008), however they are out of the scope of this article. Until now, these compressive sensing methods only could recover the gPC expansion at a particular fixed spatial input value, making them impractical if gPC expansion is needed for repeated use to obtain results at different new spatial points since these methods require one to repeat MC sampling and the compressive sensing to re-evaluate the expansion at each spatial input individually.

In this paper, we provide a fully Bayesian methodology for gPC that can evaluate the expansion as a function over both the random and spatial input domains, when the curse of dimensionality occurs. In the first instance, we provide an approximating gPC formulation where the gPC expansion is defined upon an unknown combination of significant PC bases from an available set of bases, while each PC coefficient is modeled as a series of an unknown combination of significant basis functions from another available

set of bases. In the second instance, we develop a fully Bayesian procedure called mixed shrinkage priors (MSP) that addresses the problem of specifying the composition of significant bases (gPC configuration) and evaluating the coefficients in the Bayesian model uncertainty context. We propose a Bayesian hierarchical model that involves assigning a prior probability on each gPC configuration in order to express uncertainty about significant bases, and assigning conjugate priors on the coefficients that can be thought of as mixtures of Ridge-LASSO shrinkage priors, in augmented form, able to group the bases in a desired way. We design a Gibbs sampler to address the computations in a stochastic manner. We explain how the output of the MSP procedure can be used to perform estimation, inference and prediction, in the gPC context, based on Bayesian model averaging (Madigan and Raftery, 1994; Hoeting et al., 1999) that presents optimal predictive ability, and median probability model (Barbieri and Berger, 2004) that performs dimension reduction.

The main contributions of the paper include:

- The gPC expansion can be evaluated over the whole spatial (and stochastic) domain at once even when the aforementioned curse of dimensionality issue occurs. As such, it can be re-used to obtain results at any new spatial input values without repeating MC sampling or any compressive sensing algorithm to re-evaluate the expansion at those new spatial input values.
- The proposed gPC formulation encourages the recovery of possible sparse structures in both stochastic and spatial domains.
- gPC is introduced in the Bayesian model uncertainty framework where uncertainty about the unknowns (e.g. significance of bases and associated coefficients) is expressed from the posterior distributions, allowing interval estimation, while prediction of the solution at different input values is performed via predictive posterior distributions.
- We developed the MSP procedure that accomplishes both basis selection and coefficient shrinkage simultaneously in a stochastic way, without the need of any ad hoc techniques or pilot runs.
- MSP is a fully Bayesian procedure that compromises between Bayesian Ridge and LASSO regressions (applying l_2 -norm and l_1 -norm shrinkage correspondingly) in an adaptive, data-driven manner. Combining l_2 -norm and l_1 -norm shrinkages is advantageous in estimating small but nonzero coefficients.
- MSP can consider heterogeneous sparsity in the spatial domain for different random input dimensions (or different PC coefficients), because it allows shrinkage with different combinations of l_1 -norm and l_2 -norm on different PC coefficients.

- Bayesian elastic net (BEN) (Hans, 2011) is another procedure that combines l_1 -norm and l_2 -norm shrinkages in the Bayesian framework by using a different prior specification than that of MSP. In the context of gPC, unlike BEN, MSP allows different combinations of l_1 -norm and l_2 -norm shrinkage to be applied on different PC coefficients, and hence can consider heterogeneous sparsity in the spatial domain.

This paper is organized as follows. In Section 2, we briefly review the main concepts of the gPC and describe the problem setup of the elliptic stochastic PDEs. In Section 3, we describe the proposed methodology that involves the suggested gPC formulation, and the proposed MSP procedure. In Section 4, we demonstrate the good performance of the proposed method on numerical examples with elliptic stochastic PDEs. In Section 5, we conclude and discuss further extensions and possible improvements of this work.

2. Preliminary

Briefly, we review the main concepts of the gPC expansion and set-up the problem of the elliptic SPDEs. The notation we adopt for defining vectors is based on the list comprehension syntax. For instance, $v := (a_i^{(j)}; j = 1 : 2, i = 1 : j)$ is equivalent to the 3×1 vector $v = (a_1^{(1)}, a_1^{(2)}, a_2^{(2)})$, where the former index evolves slower.

2.1. Generalized polynomial chaos

We consider a stochastic system with solution $u(x; \xi)$ that depends on a d_x -dimensional vector of spatial input variables $x \in \mathcal{X}$ and a d_ξ -dimensional vector of random input variables $\xi \in \Xi$ that admits distribution $f(\mathrm{d}\cdot)$.

The stochastic solution $u(x; \xi)$ can be represented by an infinite series of PC bases $\{\psi_\alpha(\cdot)\}$ and PC coefficients $\{g_\alpha(\cdot)\}$ in the tensor form:

$$u(x; \xi) = \sum_{\alpha \in \mathbb{N}_0^{d_\xi}} g_\alpha(x) \psi_\alpha(\xi), \quad (1)$$

for $\xi \sim f(\mathrm{d}\cdot)$ and $x \in \mathcal{X}$ (Xiu, 2010). We denote multi-indices $\alpha := (\alpha_1, \dots, \alpha_{d_\xi})$ of size d_ξ that are defined on a set of non-negative integers $\mathbb{N}_0^{d_\xi} := \{(\alpha_1, \dots, \alpha_{d_\xi}) : \alpha_j \in \mathbb{N} \cup \{0\}\}$. The family of polynomial bases $\{\psi_\alpha(\cdot); \alpha \in \mathbb{N}_0^{d_\xi}\}$ contains multidimensional orthogonal polynomial bases with respect to the probability measure $f(\mathrm{d}\cdot)$ of ξ . Each multidimensional PC basis $\psi_\alpha(\cdot)$ results as a tensor product of univariate orthogonal polynomial bases $\psi_{\alpha_j}(\cdot)$ of degree $\alpha_j \in \mathbb{N}_0^1$ namely:

$$\psi_\alpha(\xi) = \prod_{j=1}^{d_\xi} \psi_{\alpha_j}(\xi_j), \quad \alpha_j \in \mathbb{N}_0^1, \quad (2)$$

where $E_f(\psi_{\alpha_j}(\xi)\psi_{\alpha_{j'}}(\xi)) = Z_j\delta_0(j - j')$, for $j, j' = 1, \dots, d_\xi$ and $Z_j = E_f(\psi_{\alpha_j}^2(\xi))$. Here, $\delta_A(B)$ denotes the indicator function, where $\delta_A(B) = 1$ if $A = B$, and $\delta_A(B) = 0$ otherwise. It is common in practice, but not a panacea, for the family of PC bases $\{\psi_\alpha(\cdot)\}$ to be pre-defined so that they are orthogonal with respect to the distribution $f(d\cdot)$. In this way, many common distributions can be associated with a specific family of polynomials, e.g. the Askey family (Xiu and Karniadakis, 2002). Otherwise, it is possible for suitable polynomial bases to be numerically generated according to (Wan and Karniadakis, 2006). The PC coefficients $\{g_\alpha(x); \alpha \in \mathbb{N}_0^{d_\xi}\}$ can be computed as $g_\alpha(x) = E_f(u(x; \xi)\psi_\alpha(\xi))/Z_\alpha$, for $\alpha \in \mathbb{N}_0^{d_\xi}$ (Xiu, 2010), however often they are not available in closed form because the integral in the expectation is intractable.

For computational purposes, a truncated version of (1) is used by considering a finite set of available PC bases. Traditionally, the *total truncated rule* is used, which results in the expansion form:

$$u_{p_\xi}(x; \xi) := \sum_{\alpha \in A} g_\alpha(x)\psi_\alpha(\xi), \quad (3)$$

that accounts for only a finite set of multi-indices A such that $A = \{\alpha \in \mathbb{N}^{d_\xi} : \sum_{i=1}^{d_\xi} \alpha_i \leq p_\xi\}$ with cardinality $m_\xi := \frac{(p_\xi + d_\xi)!}{p_\xi! d_\xi!}$. Other truncation rules can be adopted (Blatman and Sudret, 2011; Sargsyan et al., 2013).

Theoretical results show that, for a large enough PC degree p_ξ , the truncated $u_{p_\xi}(x; \xi)$ converges to $u(x; \xi)$ under mild conditions (Xiu, 2010; Babuška et al., 2004). Following standard approximation theory and provided that $u(x; \xi)$ is square integrable with respect to $f(d\cdot)$, the gPC expansion (3) converges to $u(x; \xi)$ as:

$$\lim_{p_\xi \rightarrow \infty} E_f \left(u(x; \xi) - u_{p_\xi}(x; \xi) \right)^2 = 0,$$

and the rate of the convergence depends on the regularity of $u(x; \xi)$ (Xiu, 2010).

It is possible to describe the uncertainty of $u(x; \xi)$ with respect to ξ by computing the statistics of the gPC expansion. For instance, the solution mean $\mu(x) := E_f(u(x; \xi)) \approx E_f(u_{p_\xi}(x; \xi)) = g_0(x)$, variance $\sigma^2(x) := \text{Var}_f(u(x; \xi)) \approx \text{Var}_f(u_{p_\xi}(x; \xi)) = \sum_{\alpha \in A - \{0\}} g_\alpha^2(x) Z_\alpha$ and covariance $C(x, x') := \text{Cov}_f(u(x; \xi), u(x'; \xi)) \approx \text{Cov}_f(u_{p_\xi}(x; \xi), u_{p_\xi}(x'; \xi)) = \sum_{\alpha \in A - \{0\}} g_\alpha(x) g_\alpha(x') Z_\alpha$, where $Z_\alpha = E_f(\psi_\alpha^2(\xi))$ for $\alpha \in A$, (Xiu, 2010).

In practice, the evaluation of the gPC expansion (7) is challenging in high-dimensional scenarios if a high degree of accuracy is required. The number of unknown PC coefficients is of order $d_\xi^{p_\xi}$ and grows rapidly with the dimension d_ξ and PC degree p_ξ . This can give rise to the aforesaid curse of dimensionality and related computational problems when the training data-set is limited. Reduction of the PC degree or careless omission of PC bases in order to reduce the number of the unknowns may lead to a significant increase of bias and a poor approximation of the solution. Therefore, for the selection of significant bases, intuitive

methods able to trade off efficiently between the bias (caused by omitting bases) and the over-fitting are required.

2.2. Elliptic SPDEs with sparse solution

As discussed by Doostan and Owhadi (2011); Yang and Karniadakis (2012), elliptic SPDEs with random coefficients provide a suitable scenario for stochastic systems where the researcher can afford only a small number of system realizations due to the computationally expensive solver while the solution is sparse when expanded as a gPC.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space where \mathcal{P} is a probability measure on the σ -field \mathcal{F} . We consider the elliptic SPDE:

$$\begin{aligned} -\nabla \cdot (a(x; \omega) \nabla u(x; \omega)) &= b(x), \text{ if } x \in \mathcal{X}; \\ u(x; \omega) &= 0, \text{ if } x \in \partial\mathcal{X}, \end{aligned} \quad (4)$$

\mathcal{P} -a.s., $\omega \in \Omega$, defined on a bounded Lipschitz continuous domain $\mathcal{X} \subset \mathbb{R}^{d_x}$, $d_x = 1, 2, 3$, with boundary $\partial\mathcal{X}$.

The diffusion coefficient $a(x; \omega)$ is an unknown stochastic function defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and the source of uncertainty in (4). We model $a(x; \omega)$ as a truncated Karhunen-Loève (K-L) expansion

$$a(x; \omega) = \bar{a}(x) + \sigma_a \sum_{j=1}^{d_\xi} \sqrt{\ell_j} \varphi_j(x) \xi_j(\omega), \quad (5)$$

where $\xi := (\xi_j; j = 1 : d_\xi)$ is a d_ξ -dimensional random variable defined as $\xi : \Omega \rightarrow \Xi$ that admits distribution $f(\cdot)$, $\{(\ell_j, \varphi_j); j = 1 : d_\xi\}$ are pairs of eigenvalues and eigenfunctions of a covariance function $K_{aa}(x_1, x_2) \in L_2(\mathcal{X} \times \mathcal{X})$ of $a(x; \omega)$, $\bar{a}(x)$ is the mean of $a(x; \omega)$ and σ_a is the standard deviation that controls the variability of $a(x; \omega)$. Given that the diffusion coefficient is modeled as in (5), the solution $u(x; \omega)$ and the diffusion coefficient $a(x; \omega)$ admit representations $u(x; \xi) := u(x; \xi(\omega))$ and $a(x; \xi) := a(x; \xi(\omega))$, where $\alpha : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$ and $u : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$.

Similar to (Doostan and Owhadi, 2011; Yang and Karniadakis, 2012), we assume that $a(x; \xi)$ satisfies the following conditions (C-1, C-2, C-3):

C-1 : For all $x \in \mathcal{X}$, constants a_{\min} and a_{\max} exist, such that $0 < a_{\min} \leq a(x; \xi) \leq a_{\max} < \infty$, a.s.

C-2 : The covariance function $K_{aa}(x_1, x_2)$ is piecewise analytic on $\mathcal{X} \times \mathcal{X}$ (Bieri and Schwab, 2009; Schwab and Todor, 2006), implying that constants $c_1, c_2 \in \mathbb{R}$ exist, such that $0 \leq \ell_j \leq c_1 \exp(-c_2 j^{1/d_x})$ and $\forall a \in \mathbb{N}^{d_\xi} : \sqrt{\ell_j} \|\partial \varphi_j\|_{L^\infty(\mathcal{X})} \leq c_1 \exp(-c_2 j^{1/d_x})$ for $j = 1, \dots, d_\xi$, where $a \in \mathbb{N}$ is a multi-index.

C-3 : The random variables $\{\xi_j; j = 1 : d_\xi\}$ are independently and identically distributed according to the uniform distribution $U(-1, 1)$ in $\Xi := [-1, 1]^{d_\xi}$.

SPDE (4) can be solved by expanding the solution $u(x; \xi)$ as a gPC (3) using Legendre polynomials, where ξ is the random input variable and x is the spatial input variable. Conditions C-1, C-2 and C-3 guarantee that the solution is sparse in the stochastic domain and the gPC approximation converges exponentially fast in the mean-square sense as the PC degree increases (Babuška et al., 2004, 2007; Bieri and Schwab, 2009). Condition C-3 can be relaxed (Yang and Karniadakis, 2012).

Our interest lies in the representation of the stochastic solution $u(x; \xi)$, with respect to x and ξ , as a gPC expansion. This task can be challenging since the curse of dimensionality can occur because (i) a training data-set, with small number of training samples, is often available due to the expensive solver (ii) a huge number of unknown PC coefficients exist when ξ is large in dimension.

3. Bayesian inference of a sparse gPC expansion

We reformulate the gPC expansion defined in Section 2.1 by modeling the PC coefficients as series of basis functions. We present the proposed mixed shrinkage prior (MSP) procedure; a fully Bayesian procedure that evaluates the gPC expansion in the Bayesian model uncertainty framework and uses a Gibbs sampler to address the computations. In the Bayesian model uncertainty framework, we suggest suitable approaches to make inference and prediction with the gPC expansion, by using the output of the proposed MSP procedure.

3.1. The gPC formulation of the expansion

We consider the gPC expansion as defined in Section 2.1. In practice, the PC coefficients $\{g_a(x)\}$ are not available in closed form. We propose to model each PC coefficient $g_a(x)$ as an expansion of basis functions (herein referred to as spatial bases). We consider truncated polynomial expansions of degree p_x , in the form:

$$g_{a,p_x}(x) = \sum_{b \in B} c_{a,b} \theta_b(x), \quad x \in \mathcal{X}, \quad (6)$$

where $\{\theta_b(x); b \in B\}$ are polynomial bases and $\{c_{a,b}; a \in A, b \in B\}$ are the associated unknown coefficients $c_{a,b} \in \mathbb{R}$, here called scalar coefficients. Note that B may depend on index a ; however to simplify the presentation of the method, we suppress this indexing. Without loss of generality, when $d_x > 1$, the spatial bases can be defined as tensor products of univariate bases $\{\theta_{b_j}(\cdot); j = 1 : d_x\}$ of degree $b_j \in \mathbb{N}_0^1$ and $B = \{b \in \mathbb{N}^{d_x} : \sum_{i=1}^{d_x} b_i \leq p_x\}$, with $m_x = \text{card}(B)$, similarly to (2) and according to the total truncated

rule. Therefore, by plugging (6) in (3), we obtain $u(x; \xi) = u_{p_\xi, p_x}(x; \xi) + \varepsilon_{p_\xi, p_x}(x; \xi)$, where

$$\begin{aligned} u_{p_\xi, p_x}(x; \xi) &= \sum_{\alpha \in A} \left(\sum_{b \in B} c_{a,b} \theta_b(x) \right) \psi_\alpha(\xi) \\ &= \chi(x; \xi) c, \end{aligned} \quad (7)$$

and $\varepsilon_{p_\xi, p_x}(x; \xi)$ is the total truncation error. Here, $\chi(x; \xi) := (\theta_b(x) \psi_a(\xi); a \in A, b \in B)$ denotes the vector of basis functions and $c := (c_{a,b}; a \in A, b \in B)$ denotes the vector of scalar coefficients.

Standard polynomial and tensor product theories (Cheney, 1966; Todd, 1963) suggest that $g_{a,p_x}(x)$ converges to $g_a(x)$ as $p_x \rightarrow \infty$ under mild conditions that $g_a(\cdot)$ is smooth. It is worth noting that the spatial bases can be standard polynomial bases, e.g. Legendre, Chebyshev and Hermite (Xiu, 2010), or more sophisticated ones, e.g. B-splines bases (De Boor, 1978; Silverman, 1985; Holmes and Mallick, 2001). Here, modeling the PC coefficients as series of bases aims to facilitate the global approximation of the gPC expansion and the recovery of possible sparse structure of the solution in both the stochastic and the spatial domains.

We assume there is available a MC sample $\mathcal{D} = \left\{ (u_{i,j}, x_i^{(j)}, \xi_j) \right\}_{j=1; i=1}^{n_\xi; n_x^{(j)}}$, acting as a training data-set, where n_ξ is the number of system realizations, $n_x^{(j)}$ is the number of spatial points at j -th realization, ξ_j denotes the random input value at j -th realization, $x_i^{(j)}$ denotes the i -th spatial point at j -th realization and $u_{i,j} := u(x_i^{(j)}; \xi_j)$ denotes the solution at input values $(x_i^{(j)}, \xi_j)$. Given the MC sample \mathcal{D} and the gPC expansion (7), it is

$$u_{i,j} = u_{p_\xi, p_x}(x_i^{(j)}; \xi_j) + \epsilon_{i,j}, \text{ for } j = 1, \dots, n_\xi \text{ and } i = 1, \dots, n_x^{(j)}, \quad (8)$$

where $\epsilon_{i,j} \in \mathbb{R}$ is a residual term, associated to the (i, j) -th datum of \mathcal{D} , that refers to the discrepancy between the exact system solution and the truncated gPC expansion (7) given that the training data-set has been obtained. Therefore, residual term $\epsilon_{i,j}$ is considered to be the discrepancy for a given known-fixed input value $(x_i^{(j)}, \xi_j)$, of the (i, j) -th datum of training data-set \mathcal{D} obtained. As such, once the training data-set is obtained, given the (i, j) -th datum of \mathcal{D} , possible randomness of $\epsilon_{i,j}$ may be subject to nuisance factors; for example, related to the training data-set collection mechanism (see Section 3.2). Eq. 8 can be written as a linear system:

$$u = Xc + \epsilon, \quad (9)$$

where $u := (u_{i,j}; j = 1 : n_\xi, i = 1 : n_x^{(j)})^\top$ is the n -dimensional vector of solutions, $\epsilon := (\epsilon_{i,j}; j = 1 : n_\xi, i = 1 : n_x^{(j)})^\top$ is the n -dimensional vector of residuals, $X_{a,b} := (\psi_a(\xi_j) \theta_b(x_i^{(j)}); j = 1 : n_\xi, i = 1 : n_x^{(j)})$ is

the (a, b) -th vector of basis functions, $X := (X_{a,b}; a \in A, b \in B)$ is an $n \times m$ dimensional matrix of basis functions, $n = \sum_{j=1}^{n_\xi} n_x^{(j)}$ and $m = m_x \cdot m_\xi$.

Reformulating (3) into (9) extends the scope of existing compressive sensing methods (Doostan and Owhadi, 2011; Yang and Karniadakis, 2012; Sargsyan et al., 2013) which, until now, were able to evaluate the gPC expansion only at a fixed spatial point. By applying these compressive sensing methods to (9), one can evaluate the expansion in both stochastic and spatial domains at once and under a limited number of MC samples. In the following, we propose a new fully Bayesian procedure performing stochastic basis selection and evaluation of the gPC expansion simultaneously.

3.2. The Bayesian mixed shrinkage prior (MSP) procedure

We introduce a fully Bayesian procedure able to perform both stochastic basis selection and evaluation of the gPC expansion simultaneously. According to the standard Bayesian practice (O’Hagan and Forster, 2004), inference about the unknown quantities of a statistical model is performed based on the associated posterior distribution that is computed according to the Bayes’ theorem (Bayes and Price, 1763; de Laplace, 1812). This requires the specification of the likelihood function, and the prior model that can incorporate any available information about the unknowns before obtaining training data-set. In what follows, we consider that the gPC expansion is formulated as in (7), and that a MC sample $\mathcal{D} = \left\{ (u_{i,j}, x_i^{(j)}, \xi_j) \right\}_{j=1; i=1}^{n_\xi; n_x^{(j)}}$ (training data-set) is available.

The likelihood function $\mathcal{L}(u|c, \sigma^2) := \mathcal{L}(\{u_{i,j}\}|\{(x_i^{(j)}, \xi_j)\}, c, \sigma^2)$ is:

$$\begin{aligned} \mathcal{L}(u|c, \sigma^2) &= \prod_{j=1}^{n_\xi} \prod_{i=1}^{n_x^{(j)}} \mathbf{N} \left(u_{i,j} | \chi(x_i^{(j)}; \xi_j)^\top c, \sigma^2 \right) \\ &= \mathbf{N}_n \left(u | X^\top c, \mathbb{I}_n \sigma^2 \right), \end{aligned} \tag{10}$$

where $\mathbf{N}(\cdot|\mu, \sigma^2)$ denotes the Normal density with mean μ and variance σ^2 . The choice of the likelihood is merely a modeling choice. Here, the likelihood function can be considered as a measure of goodness-of-fit of the truncated gPC expansion to the training data-set, where the statistical discrepancy between the real model and the gPC expansion, for a given training data-set, is quantified by the residual term $\{\epsilon_{i,j}\}$. The rational is that, in the sense of maximum entropy principle (Jaynes, 2003), the residual term could be modeled approximately as a Gaussian noise, with a ‘vanishing’ mean and a variance σ^2 that represents the tolerance to the fitting error (Sargsyan et al., 2013; Cressie and Johannesson, 2008). It is worth mentioning that although the MC samples are usually generated by deterministic solvers with no random error, (Stein, 1999; Gramacy and Lee, 2012) observed that treating $\{\epsilon_{i,j}\}$ as random leads to surrogate models with better statistical

properties, possibly because it can “smooth out” several accumulated errors from: the deterministic solver, truncation of the expansion and others (Gramacy and Lee, 2012).

We define a binary vector of inclusion parameters $\gamma := (\gamma_{a,b}; a \in A, b \in B)$ that represents which of the m possible basis functions involved in the gPC expansion are significant; for example $\gamma_{a,b} = 0$ implies $c_{a,b} = 0$, while $\gamma_{a,b} = 1$ implies $c_{a,b} \neq 0$. This denotes which of those basis functions are significant or non-significant (included in or excluded from the expansion). More precisely, the PC basis $\psi_a(\cdot)$ is significant if the PC coefficient $g_a(\cdot)$ is associated to, at least, one non-zero scalar coefficient $\{c_{a,b}; b \in B\}$ or equivalently non-zero inclusion parameter.

To account for uncertainty about significant basis functions, we assign independent priors $\gamma_{a,b} \sim \text{Bernoulli}(\rho)$ with $\rho \in (0, 1)$. Regarding the scalar coefficients $c_{a,b}$, we assign priors

$$\pi(c_{a,b}|\gamma_{a,b}, \sigma^2, \eta_a, \lambda_a) = \begin{cases} \eta_a \mathcal{N}(c_{a,b}|0, \sigma^2/\lambda_a) + (1 - \eta_a) \text{DE}(c_{a,b}|0, \sigma/\lambda_a) & , \text{ if } \gamma_{a,b} = 1 \\ \delta_0(c_{a,b}) & , \text{ if } \gamma_{a,b} = 0 \end{cases},$$

where η_a is a binary shrinkage label, such that $\eta_a = 1$ indicates the Normal component (applying l_2 -norm shrinkage), while $\eta_a = 0$ indicates the Double Exponential component (applying l_1 -norm shrinkage) and $\lambda_a \in (0, \infty)$ is the shrinkage parameter for $a \in A$. Here, larger values of λ_a force $\{c_{a,b}, b \in B\}$ closer to zero. The rationale is that in order to encourage further shrinkage of small but non-zero scalar coefficients $c_{a,b}$, we assign on each $c_{a,b}$ (associated to $\gamma_{a,b} = 1$) shrinkage priors: a Normal prior that applies l_2 -norm shrinkage according to Ridge regression (Jeffrey, 1961), or a Double Exponential prior that applies l_1 -norm shrinkage according to LASSO regression (Hans, 2010; Park and Casella, 2008). These two shrinkage priors encourage different penalization and shrinkage for scalar coefficients associated to different PC coefficients by grouping them according to the PC coefficients they belong; here l_1 -norm applies a stronger shrinkage than l_2 -norm. The advantage of using two types of different shrinkage priors is that different shrinkage can be applied to different PC coefficients in order to express different levels of sparsity in spatial domain for different random input dimensions.

When m is large, the exact computation of all the posteriors $\Pr(\gamma|u, X)$ and $\pi(c|u, X, \gamma)$ is practically impossible because the number of all the different combinations of basis functions, 2^m , increases dramatically with m . For the computations, we design a Gibbs sampler (Geman and Geman, 1984) that generates a Markov chain which visits different gPC configurations proportionally to $\Pr(\gamma|u, X)$. Thus, the procedure spends most of the evaluation time at important (highly possible) gPC configurations. We consider a Gibbs sweep with blocks updating each $\{(\gamma_{a,b}, c_{a,b}); a \in A, b \in B\}$ recursively. Direct sampling from the full

conditional posterior of $(\gamma_{a,b}, c_{a,b})$ is possible for $a \in A, b \in B$. This can be performed by decomposing the associated conditional posterior distribution into:

$$\begin{aligned} \pi(\gamma_{a,b}, c_{a,b} | u, X, \gamma_{-(a,b)}, c_{-(a,b)}, \eta_a, \rho, \sigma^2, \lambda_a) &= \Pr(\gamma_{a,b} | u, X, \gamma_{-(a,b)}, c_{-(a,b)}, \eta_a, \rho, \sigma^2, \lambda_a) \\ &\quad \times \pi(c_{a,b} | u, X, \gamma, c_{-(a,b)}, \eta_a, \rho, \sigma^2, \lambda_a), \end{aligned} \quad (11)$$

for $a \in A, b \in B$. Here, $X_{-(a,b)}$ and $c_{-(a,b)}$ denotes the sub-matrix of X and sub-vector of c that result after removing the (a,b) column and element correspondingly. The conditional probability of $\gamma_{a,b}$ on the right-hand side of (11) is $\text{Bernoulli}(P_{a,b}^{(\gamma)})$ with acceptance probability:

$$P_{a,b}^{(\gamma)} = \begin{cases} \left[1 + 2 \frac{1-\rho}{\rho} \frac{\sigma}{\lambda_a} / \left(\frac{\Phi(-\mu_{a,b}^-/s_{a,b})}{\mathbf{N}(0|\mu_{a,b}^-, s_{a,b}^2)} + \frac{\Phi(\mu_{a,b}^+/s_{a,b})}{\mathbf{N}(0|\mu_{a,b}^+, s_{a,b}^2)} \right) \right]^{-1}, & \eta_a = 0 \\ \left[1 + \frac{1-\rho}{\rho} \sqrt{\frac{2\pi\sigma^2}{\lambda_a}} \mathbf{N}(0|\mu_{a,b}^*, s_{a,b}^{2,*}) \right]^{-1}, & \eta_a = 1 \end{cases}, \quad (12)$$

and

$$\mu_{a,b}^+ = \left(X_{a,b}^\top X_{a,b} \right)^{-1} \left[X_{a,b}^\top (u - X_{-(a,b)} c_{-(a,b)}) - \sigma \lambda_a \right]; \quad (13)$$

$$= \left(X_{a,b}^\top X_{a,b} \right)^{-1} \left[X_{a,b}^\top u - X_{a,b}^\top X_{-(a,b)} c_{-(a,b)} - \sigma \lambda_a \right];$$

$$\mu_{a,b}^- = \left(X_{a,b}^\top X_{a,b} \right)^{-1} \left[X_{a,b}^\top (u - X_{-(a,b)} c_{-(a,b)}) + \sigma \lambda_a \right]; \quad (14)$$

$$= \left(X_{a,b}^\top X_{a,b} \right)^{-1} \left[X_{a,b}^\top u - X_{a,b}^\top X_{-(a,b)} c_{-(a,b)} + \sigma \lambda_a \right];$$

$$s_{a,b}^2 = \sigma^2 (X_{a,b}^\top X_{a,b})^{-1}; \quad (15)$$

$$\mu_{a,b}^* = \left(X_{a,b}^\top X_{a,b} + \lambda_a \right)^{-1} X_{a,b}^\top (u - X_{-(a,b)} c_{-(a,b)}); \quad (16)$$

$$= \left(X_{a,b}^\top X_{a,b} + \lambda_a \right)^{-1} \left[X_{a,b}^\top u - X_{a,b}^\top X_{-(a,b)} c_{-(a,b)} \right];$$

$$s_{a,b}^{2,*} = \sigma^2 (X_{a,b}^\top X_{a,b} + \lambda_a)^{-1}, \quad (17)$$

for $a \in A, b \in B$. The full conditional posterior distribution of $c_{a,b}$ in (11) admits density:

$$\pi(c_{a,b} | u, X, \gamma, c_{-(a,b)}, \eta_a, \rho, \sigma^2, \lambda_a) = \begin{cases} \delta_{\{0\}}(c_{a,b}) & , \text{ if } \gamma_{a,b} = 0 \\ \mathbf{N}(c_{a,b} | \mu_{a,b}^*, s_{a,b}^{2,*}) & , \text{ if } \gamma_{a,b} = 1, \eta_a = 1 \\ (1 - w_{a,b}) \mathbf{N}^-(c_{a,b} | \mu_{a,b}^-, s_{a,b}^2) \\ \quad + w_{a,b} \mathbf{N}^+(c_{a,b} | \mu_{a,b}^+, s_{a,b}^2) & , \text{ if } \gamma_{a,b} = 1, \eta_a = 0 \end{cases}, \quad (18)$$

where

$$w_{a,b} = \left[1 + \frac{\Phi(-\mu_{a,b}^-/s_{a,b}) \mathbf{N}(0|\mu_{a,b}^+, s_{a,b}^2)}{\mathbf{N}(0|\mu_{a,b}^-, s_{a,b}^2) \Phi(\mu_{a,b}^+/s_{a,b})} \right]^{-1}, \quad (19)$$

for $a \in A, b \in B$. Here, $\Phi(\cdot)$ denotes the cumulative density function of the standard normal distribution, while $\mathbf{N}^-(x|\mu, s^2) := \frac{\mathbf{N}(x|\mu, s^2)}{\Phi(-\mu/s)} \delta_{(-\infty, 0)}(x)$ and $\mathbf{N}^+(x|\mu, s^2) := \frac{\mathbf{N}(x|\mu, s^2)}{\Phi(\mu/s)} \delta_{(0, +\infty)}(x)$ denote the right and left truncated Normal distribution densities. A rough sketch on the derivation of the of Eq. (12)-(19) is given in the Appendix. In what follows, we add progressively more blocks to the Gibbs sweep as we consider more levels of uncertainty.

Different values of σ^2 represent different amounts of residual variation and thus setting σ^2 fixed may lead to misleading results when different subsets of basis functions are considered. We treat σ^2 as unknown and assign a prior $\sigma^2 \sim \mathbf{G}(a_\sigma, b_\sigma)$ with $a_\sigma > 0$ and $b_\sigma > 0$. Choosing Gamma priors for the residual variance is a standard practice in Bayesian regression (O'Hagan and Forster, 2004). Weakly informative priors can be considered by letting $a_\sigma \rightarrow 0$ and $b_\sigma \rightarrow 0$. The resulting full conditional posterior distribution is not of a standard form and has density known up to a normalizing constant as:

$$\begin{aligned} \pi(\sigma^2|u, X, \gamma, c, \eta, \lambda) \propto & \left(\frac{1}{\sigma} \right)^{\frac{n}{2} + \frac{m\gamma}{2} + a_\sigma + 1} \exp \left(-\frac{1}{2\sigma^2} |u - Xc|_2^2 \right. \\ & \left. - \frac{1}{2\sigma^2} \sum_{\{a \in A: \eta_a = 1\}} \lambda_a |c_a|_2^2 - \frac{1}{\sigma} \sum_{\{a \in A: \eta_a = 0\}} \lambda_a |c_a|_1 - \frac{b_\sigma}{\sigma^2} \right), \end{aligned} \quad (20)$$

where $c_a := (c_{a,b}; a \in A)$. However, it can be sampled exactly by using adaptive rejection sampling (ARS) (Gilks, 1992; Gilks and Wild, 1992) since (20) is log-concave. Alternatively, one can employ a random walk Metropolis (RWM) update (Metropolis et al., 1953; Robert and Casella, 2004) where the scale of the proposal is properly calibrated by an adaptation scheme (Andrieu and Thoms, 2008) that forces the expected acceptance probability to be close to its optimal value 0.5 (Roberts et al., 1997).

To account for uncertainty about the shrinkage norm, controlled by the vector of shrinkage labels $\eta := (\eta_a; a \in A)$, and shrinkage parameter $\lambda := (\lambda_a; a \in A)$, we assign priors $\eta_a|\varpi \sim \text{Bernoulli}(\varpi)$ with hyperparameter $\varpi \in (0, 1)$ and $\lambda_a \sim \mathbf{G}(a_\lambda, b_\lambda)$ with parameters $a_\lambda > 0$ and $b_\lambda > 0$ for $a \in A$. Gamma distribution is a convenient choice of priors for the shrinkage parameters mainly because they are conjugate and there are theoretical guidelines on how to choose the associated hyper-parameters (Lykou and Ntzoufras, 2012). It is worth noting that marginalizing the conditional prior $\pi(c_{a,b}, \eta_a|\gamma_{a,b} = 1, \varpi, \lambda_a, \sigma^2) = \pi(c_{a,b}|\gamma_{a,b} = 1, \eta_a, \lambda_a, \sigma^2) \Pr(\eta_a|\varpi)$ over the shrinkage labels η_a results a mixture of Normal-Double Exponential shrinkage priors $\pi(c_{a,b}|\gamma_{a,b} = 1, \varpi, \lambda_a, \sigma^2) = \varpi \mathbf{N}(c_{a,b}|0, \sigma^2/\lambda_a) + (1 - \varpi) \text{DE}(c_{a,b}|0, \sigma/\lambda_a)$, that apply a combination of

l_2 -norm and l_1 -norm shrinkages. As noticed by Hans (2011), combining l_2 -norm and l_1 -norm shrinkages is advantageous in estimating small but nonzero coefficients. Direct sampling from the full conditional posterior of (η_a, λ_a) is possible for $a \in A$, as:

$$\pi(\eta_a, \lambda_a | u, X, \gamma_a, c_a, \sigma^2, \varpi) = \Pr(\eta_a | u, X, \gamma_a, c_a, \sigma^2, \varpi) \pi(\lambda_a | u, X, \gamma_a, c_a, \eta_a, \sigma^2),$$

where the conditional posterior probability of η_a on the right-hand side is a **Bernoulli**($P_a^{(\eta)}$) with acceptance probability

$$P_a^{(\eta)} = \left[1 + \frac{C_a^{(0)}}{C_a^{(1)}} \frac{1 - \varpi}{\varpi} (\pi/2)^{m_{\gamma|a}/2} \right]^{-1}, \quad (21)$$

where $C_a^{(0)} = \Gamma(m_{\gamma|a} + a_\lambda) \left(\frac{|c_a|_1}{\sigma} + b_\lambda \right)^{-(m_{\gamma|a} + a_\lambda)}$ and $C_a^{(1)} = \Gamma(m_{\gamma|a}/2 + a_\lambda) \left(\frac{|c_a|_2^2}{2\sigma^2} + b_\lambda \right)^{-(m_{\gamma|a}/2 + a_\lambda)}$, while that of λ_a on the left-hand side is a Gamma $\lambda_a | \gamma_a, c_a, \sigma^2, \eta_a \sim \eta_a \mathcal{G} \left(\frac{m_{\gamma|a}}{2} + a_\lambda, \frac{1}{2\sigma^2} |c_a|_2^2 + b_\lambda \right) + (1 - \eta_a) \mathcal{G} \left(m_{\gamma|a} + a_\lambda, \frac{1}{\sigma} |c_a|_1 + b_\lambda \right)$. Of note, $\{(\eta_a, \lambda_a); a \in A\}$ are aposteriori conditionally independent. Effectively, integrating over $\{(\eta_a, \lambda_a)\}$, the procedure creates aposteriori a mixture of Ridge-LASSO regression models, allowing for adaptive, data-based shrinkage of the scalar coefficients. Compared to other non-Bayesian approaches, the proposed procedure avoids the need for ad hoc techniques, such as Cross-Validation, or pilot runs and allows the data to determine suitable values for the shrinkage quantities.

To account for uncertainty about the hyper-priors ρ and ϖ , we assign Beta priors $\rho | a_\rho, b_\rho \sim \mathbf{Beta}(a_\rho, b_\rho)$ with $a_\rho > 0$, $b_\rho > 0$ and $\varpi | a_\varpi, b_\varpi \sim \mathbf{Beta}(a_\varpi, b_\varpi)$ with $a_\varpi > 0$, $b_\varpi > 0$. Moreover, non-informative priors can be considered, if $a_\rho = b_\rho = 1$, for ρ and if $a_\varpi = b_\varpi = 1$, for ϖ . We chose Beta priors for practical reasons because they are conjugate. The hyper-priors ρ and ϖ are aposteriori independent with full conditional posteriors $\rho | \gamma \sim \mathbf{Beta}(m_\gamma + a_\rho, m - m_\gamma + b_\rho)$ and $\varpi | \eta \sim \mathbf{Beta}(m_\eta + a_\varpi, m_\xi - m_\eta + b_\varpi)$, where $m_\eta = \sum_{a \in A} \eta_a$. The importance of treating the shrinkage hyper-parameter ρ as unknown has been noticed by Scott and Berger (2010) who observed that even for seemingly reasonable choices of fixed values, conditioning on ρ may lead to unsatisfactory results.

The pseudo-code of the Gibbs sweep is summarized in Alg. 1. The updates within Blocks III and IV can be performed in parallel to reduce the execution time of the code. We refer to the proposed procedure as mixed shrinkage prior (MSP), highlighting a main ingredient of the method.

Remark 1. *The proposed method admits, as limiting cases, the Bayesian LASSO (Hans, 2010) using a strong l_1 -norm shrinkage via Double Exponential priors for $\varpi \rightarrow 0$ and the Ridge regression (Jeffrey, 1961) using a strong l_2 -norm shrinkage via Normal distribution prior for $\varpi \rightarrow 1$. By treating ϖ as unknown, the procedure can adjust the shrinkage method between l_1 - and l_2 -norms for each PC coefficient in an adaptive, data-based manner.*

Remark 2. *Increasing the size of the training data-set (n_x , n_ξ and n) does not increase the computational*

Algorithm 1 Blocks of the Gibbs sweep

Block I Update $\{(\gamma_{a,b}, c_{a,b})\}$: For $a \in A, b \in B$,

1. Compute $\mu_{a,b}^-, \mu_{a,b}^+, s_{a,b}^2, \mu_{a,b}^*, s_{a,b}^{2,*}$ according to (13), (14), (15), (16) and (17).
2. Update $\gamma_{a,b}$: draw $\gamma_{a,b}$ from $\text{Bernoulli}(P_{a,b}^{(\gamma)})$, according to (12).
3. Update $c_{a,b}$: draw $c_{a,b}$ from $\pi(c_{a,b}|u, X, \gamma, c_{-(a,b)}, \eta_a, \rho, \sigma^2, \lambda_a, \varpi)$ according to (18).

Block II Update σ^2 : Draw σ^2 from ARS (or RWM update with adaptation).

Block III Update (η_a, λ_a) : For $a \in A$,

1. Draw η_a from $\text{Bernoulli}(P_a^{(\eta)})$, according to (21).
2. Draw λ_a from $\eta_a \mathcal{G}\left(\frac{m_{\gamma|a}}{2} + a_{\lambda}, \frac{1}{2\sigma^2} |c_a|_2^2 + b_{\lambda}\right) + (1 - \eta_a) \mathcal{G}\left(m_{\gamma|a} + a_{\lambda}, \frac{1}{\sigma} |c_a|_1 + b_{\lambda}\right)$.

Block IV Update (ρ, ϖ) :

- Draw ρ from $\text{Beta}(m_{\gamma} + a_{\rho}, m - m_{\gamma} + b_{\rho})$.
 - Draw ϖ from $\text{Beta}(m_{\eta} + a_{\varpi}, m_{\xi} - m_{\eta} + b_{\varpi})$.
-

cost of running Alg.1, significantly. The reason is because the quantities $u^{\top}u$, $X^{\top}X$ and $X^{\top}u$, whose computational cost increases with n_x , n_{ξ} and n , can be computed offline and before the execution of the algorithm. Therefore, in practice, one need to compute the quantities $u^{\top}u$, $X^{\top}X$ and $X^{\top}u$ only once before the execution of Alg. 1, so that large values of n_x , n_{ξ} and n cannot increase significantly the computational cost.

3.3. Evaluation

Given the likelihood model in (10), we can consider an estimator for the stochastic solution as:

$$u_{p_x, p_{\xi}}^{(\text{gPC})}(x; \xi) = \mathbb{E}_{\mathcal{L}}(u|x; \xi) = \sum_{a \in A} \sum_{b \in B} c_{a,b} \theta_b(x) \psi_a(\xi). \quad (22)$$

Consequently¹, the solution mean, $\mu(x)$, is evaluated as:

$$\mu^{(\text{gPC})}(x) = \mathbb{E}_f \mathbb{E}_{\mathcal{L}}(u|x; \xi) = \sum_{b \in B} c_{0,b} \theta_b(x), \quad (23)$$

the solution variance, $\sigma^2(x)$, is evaluated as:

$$\sigma^{2,(\text{gPC})}(x) = \text{Var}_f \mathbb{E}_{\mathcal{L}}(u|x; \xi) + \mathbb{E}_f \text{Var}_{\mathcal{L}}(u|x; \xi) = \sum_{a \in A - \{0\}} \left(\sum_{b \in B} c_{a,b} \theta_b(x) \right)^2 Z_a + \sigma^2, \quad (24)$$

¹Based on equations $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$, $\text{Var}(X) = \text{Var}(\mathbb{E}(X|Y)) + \mathbb{E}(\text{Var}(X|Y))$ and $\text{Cov}(X, X') = \text{Cov}(\mathbb{E}(X|Y), \mathbb{E}(X'|Y)) + \mathbb{E}(\text{Cov}(X, X')|Y)$, (Casella and Berger, 1990)

and the solution covariance, $C(x, x')$, is evaluated as:

$$\begin{aligned} C^{(\text{gPC})}(x, x') &= \text{Cov}_f(\mathbb{E}_{\mathcal{L}}(u|x; \xi), \mathbb{E}_{\mathcal{L}}(u'|x'; \xi)) + \mathbb{E}_f(\text{Cov}_{\mathcal{L}}(u, u'|x, x'; \xi)) \\ &= \sum_{a \in A - \{0\}} \left(\sum_{b \in B} c_{a,b} \theta_b(x) \right) \left(\sum_{b' \in B} c_{a,b'} \theta_{b'}(x') \right) Z_a + \delta_0(x - x') \sigma^2, \end{aligned} \quad (25)$$

where $u = \chi(x; \xi) + \epsilon$ and $u' = \chi(x'; \xi) + \epsilon'$ according to the formalism in (8 & 10). Other statistical quantities of the solution can be evaluated likewise. We observe that under the proposed setting, the estimator of $\sigma^2(x)$, (24 & 25), takes into account the residual variance σ^2 , unlike the traditional estimation practice in gPC literature that ignores it. This aims to provide more reliable estimates for $\sigma^2(x)$ since the residual variance includes the contribution of the truncated terms of the gPC expansion. Yet, being fully Bayesian, the MSP procedure can compute the point estimate, and whole posterior density, of σ^2 directly.

The proposed method is fully Bayesian. As such, inference about all the unknown quantities of the gPC expansion can be performed via the associated posterior distributions while prediction about solutions at new input values given the training data-set is performed via the predictive distribution. In what follows, we describe how the output of the MSP procedure can be used to perform estimation, inference and prediction with the truncated gPC expansion in the Bayesian framework. We consider two fully Bayesian approaches: the Bayesian model averaging (Hoeting et al., 1999) and the median probability model (Barbieri and Berger, 2004).

3.3.1. The Bayesian model average based evaluation

The evaluation of the expansion can be performed by Bayesian model averaging (BMA) (Hoeting et al., 1999) if the predictive ability of the gPC expansion is of main interest. In the gPC context, BMA accounts for uncertainty about the significant basis functions by combining inferences from different gPC configurations and weighting them according to the associated posterior inclusion probabilities. Compared to other methods that assume a single gPC configuration by selecting a fixed subset of possibly significant basis functions, BMA presents optimal predictive performance (Hoeting et al., 1999; Madigan and Raftery, 1994).

We consider a Gibbs sample $\{(\gamma^{(t)}, c^{(t)}, \sigma^{2,(t)}, \dots)\}_{t=1}^T$ generated by Alg. 1. Estimates and associated standard errors of σ^2 and $\{c_{a,b}\}$ can be derived by the ergodic averages according to the standard Bayesian practice (Robert and Casella, 2004), e.g. $\hat{c}_{a,b} = \frac{1}{T} \sum_{t=1}^T c_{a,b}^{(t)}$ and $\text{s.e.}(\hat{c}_{a,b}) = \sqrt{s_{a,b}^{c,2} \sqrt{\frac{\varrho_{a,b}^c}{T}}}$ where $s_{a,b}^{c,2}$ is the sample variance and $\varrho_{a,b}^c$ is the integrated autocorrelation time of $\{c_{a,b}^{(t)}; t = 1 : T\}$ for $a \in A$, $b \in B$. For $\mu^{(\text{gPC})}(x)$, $\sigma^{2,(\text{gPC})}(x)$, $C(x, x')$, $g_a(x)$ and $u_{p_x, p_\xi}^{(\text{gPC})}(x; \xi)$, point estimates can be computed from the Gibbs sample analogously, e.g. $\hat{\sigma}^{2,(\text{gPC})}(x) = \frac{1}{T} \sum_{t=1}^T [\sum_{a \in A - \{0\}} \left(\sum_{b \in B} c_{a,b}^{(t)} \theta_b(x) \right)^2 Z_a + \sigma^{2,(t)}]$.

It is worth noting that estimates for $\mu^{(\text{gPC})}(x)$, $g_a(x)$ and $u^{(\text{gPC})}(x; \xi)$ can be computed equivalently by using the sufficient statistics $\hat{c} := (\hat{c}_{a,b}; a \in A, b \in B)$ in equations (23), (6) and (22) in order to reduce the memory storage requirements when applied in practice. $q\%$ -Interval estimation can be performed via $q\%$ -credible intervals (CI); that is the set of values bounded by the contour of the associated posterior distribution for a fixed volume $q/100$. Prediction of the solution $u' := u(x = x'; \xi = \xi')$ at given new input values $(x'; \xi')$ and conditional on a training data-set \mathcal{D} can be performed via the predictive posterior distribution $\mathcal{L}(u'|\mathcal{D}) = \int \mathcal{L}(u'|c, \sigma^2) \pi(c, \sigma^2|\mathcal{D}) d(c, \sigma^2)$ which is approximated numerically by $\hat{\mathcal{L}}(u'|\mathcal{D}) = \frac{1}{T} \sum_{t=1}^T \mathbf{N}(u'|u_{p_x, p_\xi}^{(t)}(x'; \xi'), \sigma^{2, (t)})$, where $u_{p_x, p_\xi}^{(t)}(x; \xi) = \chi(x; \xi) c^{(t)}$.

3.3.2. Median probability model based evaluation

The gPC expansion may need to be evaluated based on a single combination of significant basis functions if there is interest in recovering a sparse (or parsimonious) representation of the stochastic solution. A suitable probabilistic basis selection mechanism is that of the median probability model (MPM) (Barbieri and Berger, 2004), which is mostly used for variable selection in Bayesian regression models. Barbieri and Berger (2004) provide theoretical results showing that MPM leads to optimal Bayesian predictive model selection (namely close to BMA) under conditions, and suggest that MPM may be successful even when these conditions are not met. In the gPC context, MPM refers to the gPC that includes only the basis functions associated with marginal inclusion posterior probabilities greater than 0.5.

Given a Gibbs sample $\{(\gamma^{(t)}, c^{(t)}, \sigma^{2, (t)}, \dots)\}_{t=1}^T$ drawn from Alg. 1, the marginal inclusion posterior probabilities can be estimated as $\hat{P}_{a,b} = \frac{1}{T} \sum_{t=1}^T \gamma_{a,b}^{(t)}$. According to the MPM rule, the inclusion parameters are estimated as $\hat{\gamma}_{a,b}^{(\text{MPM})} = \delta_{(0.5, 1)}(\hat{P}_{a,b})$, for $a \in A, b \in B$. It is reasonable to consider that significant scalar coefficients $\{c_{a,b}\}$ are those that correspond to non-zero inclusion parameters $\{\gamma_{a,b}\}$, significant PC coefficients $\{g_a(\cdot)\}$ are those that have at least one significant scalar coefficient $\{c_{a,b}\}$ and thus significant PC bases $\{\psi_a(\cdot)\}$ are those that correspond to significant PC coefficients $\{g_a(\cdot)\}$. This allows the recovery of possible sparse structures in both stochastic and spatial domains, without loosing in accuracy.

For a chosen gPC configuration according to the MPM rule, inference about the unknown quantities of interest can be performed using the part of Gibbs sample which corresponds to the significant subset of basis functions, namely $\{t = 1, \dots, T : \gamma^{(t)} = \hat{\gamma}^{(\text{MPM})}\}$. In practice, however, it is preferable to increase the size of the available Gibbs sample up to a desired number, e.g. T , by re-running the Gibbs sampler (Alg. 1) for fixed $\gamma^{(t)} = \hat{\gamma}^{(\text{MPM})}$. Then, given the new Gibbs sample, inference about unknown quantities of interest (e.g. solution mean) or prediction about solutions at new random or spatial input values can be performed as described in Section 3.3.1.

MPM evaluation primarily is recommended when the solution is sparse or a parsimonious representation of the solution is preferred. Running Alg. 1 is cheaper if the solution is sparse because non-significant scalar coefficients are omitted from the simulations, and hence the algorithm can perform fewer operations in practice.

4. Numerical examples

We assume that interest lies in the representation of the solution $u(x; \xi)$ of the 1D elliptic SPDE:

$$\begin{aligned} \frac{d}{dx} \left(a(x; \xi) \frac{d}{dx} u(x; \xi) \right) &= -1, \text{ if } \xi \in \Xi, x \in \mathcal{X}; \\ u(0; \xi) &= 0, u(1; \xi) = 0, \text{ if } \xi \in \Xi, \end{aligned} \quad (26)$$

where $\mathcal{X} = (0, 1)$, $\Xi = [-1, 1]^{d_\xi}$ and $d_\xi \in \mathbb{N}^+$, as a function of the random and spatial input variables. Here, the solution $u(x; \xi)$ is expanded as a finite gPC using Legendre polynomial bases. The associated PC coefficients are modeled as a series of Legendre bases scaled linearly in $(0, 1)$. We setup the Bayesian hierarchical model with hyper-parameters $a_\lambda = 10^{-3}$, $b_\lambda = 10^{-3}$, $a_\sigma = 10^{-3}$, $b_\sigma = 10^{-3}$, $a_\rho = 1$, $b_\rho = 1$, $a_\varpi = 1$ and $b_\varpi = 1$ that lead to weakly informative priors and evaluate the expansion by the proposed procedure.

The proposed Bayesian method is demonstrated on a tractable 1-dimensional scenario, while its performance is tested on two intractable high-dimensional scenarios with 14 and 40 dimensions in random input. Under the proposed gPC formulation, we compare the MSP procedure with the Bayesian compressive sensing (BCS) (Ji et al., 2008), l_1 -minimization (l_1 -min) (Doostan and Owhadi, 2011) and Bayesian elastic net (BEN) (Hans, 2011) by applying them on the linear system (9). BCS and l_1 -min have been recently proposed in the gPC context by Sargsyan et al. (2013) and Doostan and Owhadi (2011); Yang and Karniadakis (2012) for evaluating the gPC expansion at only a fixed spatial point, while BEN, another Bayesian procedure, is used for the first time in the gPC context here.

As measures of performance, we use the relative error of the solution mean $\epsilon(\mu; x) := \left| 1 - \frac{\hat{\mu}(x)}{\mu(x)} \right|$ and standard deviation $\epsilon(\sigma; x) := \left| 1 - \frac{\hat{\sigma}(x)}{\sigma(x)} \right|$, following Doostan and Owhadi (2011); Yang and Karniadakis (2012). Given a test data-set $\mathcal{D}' = \{(u'_{i,j}, x'_i, \xi'_j)\}_{j=1; i=1}^{n'_\xi; n'_x}$, for $n'_x = 101$ and $n'_\xi = 200$, the performance is measured globally by the total relative error of the solution mean $\epsilon(\mu) = \frac{1}{n'_x} \sum_{i=1}^{n'_x} \epsilon(\mu; x'_i)$, standard deviation $\epsilon(\sigma) = \frac{1}{n'_x} \sum_{i=1}^{n'_x} \epsilon(\sigma; x'_i)$ and mean squared predictive error $\text{mspe}(u) = \frac{1}{n'_x} \sum_{i=1}^{n'_x} \frac{1}{n'_\xi} \sum_{j=1}^{n'_\xi} (u'_{i,j} - \hat{u}_{p_x, p_\xi}(x'_i; \xi'_j))^2$.

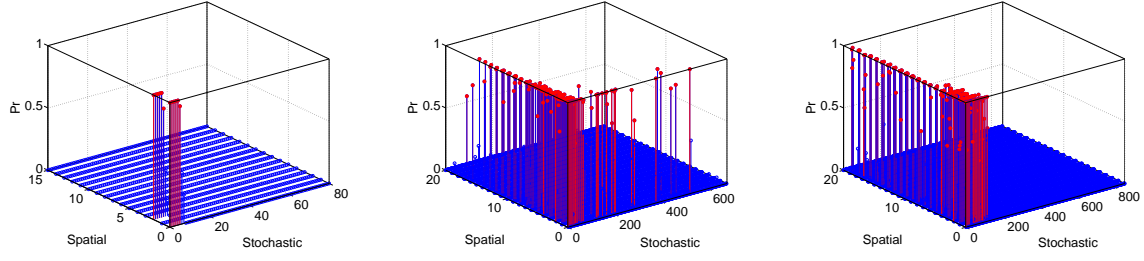
4.1. One-dimensional example: case [1D]

The diffusion coefficient in SPDE (26) is assumed to be $a(x; \xi) = a(\xi) = 1 + 0.5\xi$, where ξ is distributed according to $U(-1, 1)$ and $d_x = 1$. This leads to a tractable exact solution $u(x; \xi) = \frac{x(1-x)}{2+\xi}$ with mean $\mu(x) = x(1-x)\frac{\log(3)}{2}$, standard deviation $\sigma(x) = x(1-x)\sqrt{\frac{1}{3} - \frac{1}{4}\log^2(3)}$ and covariance $C(x, x') = x(1-x)x'(1-x')(\frac{1}{3} - \frac{1}{4}\log^2(3))$, for $x, x' \in (0, 1)$, (see appendix in (Karagiannis and Lin, 2014)).

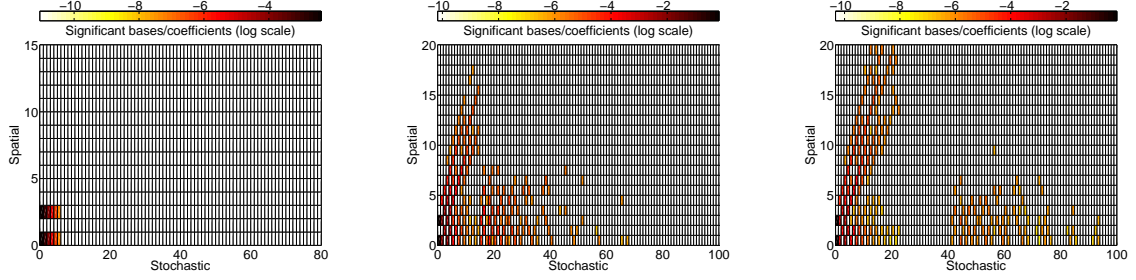
We wish to represent the solution of the SPDE as a gPC expansion with PC degree $p_\xi = 80$ and $p_x = 15$. However, a MC sample of only $n_x = 5$ and $n_\xi = 20$ system realizations is assumed to be available. Here, the solution is sparse in stochastic domain because C-I, C-II & C-III are satisfied, and in spatial domain because the solution is a 2-nd order polynomial with respect to x . We apply the MSP procedure where the Gibbs sampler (Alg. 1) runs for $2 \cdot 10^5$ iterations, of which the first 10^5 are discarded as burn-in.

The inclusion posterior probabilities are presented in Fig. 1a and the significant scalar coefficients, selected according to the MPM rule, are presented in Fig. 1d. We observe that the proposed method has recovered a sparse representation both in stochastic and spatial domains. According to the MPM rule, 6 significant PC coefficients were selected. For each significant PC coefficient, 2 spatial bases, the 0-th and the 2-nd, were selected as significant. These two spatial bases can recover the solution completely in the spatial domain because the solution is a 2-nd degree polynomial with respect to x . Therefore, in this case, the proposed method –based on the MPM rule– managed to select a minimal subset of significant spatial bases. In Fig. 1d, we observe that the absolute values of the scalar coefficients reduce with the order of the associated PC bases.

The estimation of the solution $u(x; \xi)$ as functions of x and ξ , and the associated standard error, are represented in Figs. 2a & 2b. We observe that the stochastic solution is satisfactorily recovered with standard error small in value ($< 2e - 6$) and smooth over the input domain. The proposed method is able to produce ‘interval’ estimates and predictions. The predictive distribution of the solution $u' := u(x = 0.5; \xi = \xi')$ at given new random input values of interest in the range $\xi' \in (-1, 1)$ and spatial input value at $x' = 0.5$, is presented in Fig. 2c in log scale. We observe that the predictive density of the solution is quite narrow (Fig. 2c) indicating small predictive error at given input values ($x = 0.5; \xi = \xi'$) of interest. Figs. 3a, 3j, 3g & 3d plot the estimates and associated 95%-CI for the solution statistics and PC coefficients as functions of x . So, the proposed Bayesian method can quantify uncertainty about the expansion itself as well, compared to existing compressive sensing methods in gPC (Doostan and Owhadi, 2011; Yang and Karniadakis, 2012). No visual difference regarding the estimates is observed between the results obtained by MSP-BMA and MSP-MPM since the corresponding lines are very close to each other.



(a) [1D] Inclusion posterior probabilities (b) [14D] Inclusion posterior probabilities (c) [40D] Inclusion posterior probabilities



(d) [1D] Significant scalar coefficients (e) [14D] Significant scalar coefficients (f) [40D] Significant scalar coefficients

Figure 1: Plots of the marginal inclusion posterior probabilities (panels a-c) and the significant scalar coefficients (in log scale) (panels d-f) arranged according to the indexes in the stochastic and spatial domains. The color in (panels d-f) denotes the absolute value of the scalar coefficients in log scale. Here, log denotes the natural logarithm.

Figs. 4a & 4b offer a comparison of the proposed MSP procedure against BEN, l_1 -min and BCS with respect to the relative errors of the solution mean and standard deviation as functions of x . We observe that the relative errors of MSP and BEN are smooth and constant with x while those of l_1 -min and BCS present some bumps. This shows that the former methods, which perform a probabilistic basis selection, have selected more suitable spatial bases (0-th and 2-nd), as significant, and more accurately estimated the coefficients than the latter ones, which mainly shrink all of the coefficients toward zero. In this example, we observe that under the proposed gPC formulation, MSP and BEN provide results with similar relative errors uniformly over the spatial domain.

Figs. 5a, 5d & 5g present the total relative errors of the solution statistics and mspe. The basis selection approaches (MSP, BEN) presented better performance than the compressive sensing approaches l_1 -min and BCS, regarding the $\epsilon(\mu)$. We observe that the $\epsilon(\sigma)$ of MSP is slightly smaller than that of BCS. In this simple problem, the total relative errors of MSP and BEN are be close in value. We observe that, in the gPC context, the MSP-MPM results in small relative errors and mpse that are close in value to those of MSP-BMA. This indicates that the MPM-based evaluation is suitable for recovering a sparse representation

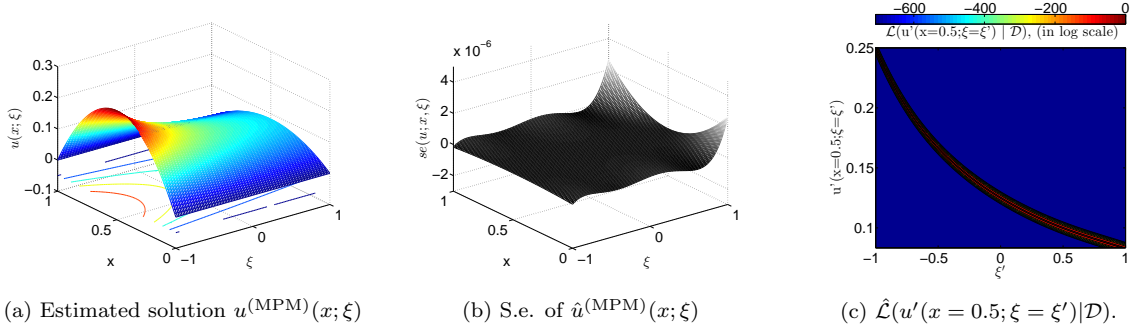


Figure 2: [1D] The estimated solution (panel a) and the associated standard error (panel b) as functions of x and ξ , for $a(x; \xi) = 1 + 0.5\xi$. The estimated predictive distribution (panel c), in log scale, of a solution $u'(x = 0.5; \xi = \xi')$ at given random input values in the range $\xi' \in (-1, 1)$ and spatial input value at $x' = 0.5$. The results were produced using the MSP procedure after iterating Alg. 1 for 10^5 times and evaluation was based on MPM. Here, log denotes the natural logarithm.

of the solution in both stochastic and spatial domains without a significant loss in accuracy.

4.2. High-dimensional example: cases [14D] & [40D]

We assume that the diffusion coefficient $\alpha(x; \xi)$ is not available in closed form but modeled by a K-L expansion that uses the d_ξ largest eigenvalues and eigenfunctions of covariance kernel $K_{aa}(x_1, x_2) = \exp\left(-\frac{(x_1 - x_2)^2}{l_K^2}\right)$, where $l_K \in (0, +\infty)$ is the correlation length that dictates the decay of the spectrum of $\alpha(x; \xi)$.

We collect a MC sample by performing system realizations of the SPDE (26) that use the deterministic solver in (Yang and Karniadakis, 2012). Herein, we consider two high-dimensional scenarios: [14D] with $d_\xi = 14$, $l_K = 1/5$, $\bar{\alpha} = 0.1$, $\sigma = 0.03$, $p_\xi = 3$, $p_x = 20$, $n_x = 11$ and $n_\xi = 120$ and [40D] with $d_\xi = 40$, $l_K = 1/14$, $\bar{\alpha} = 0.1$, $\sigma = 0.021$, $p_\xi = 2$, $p_x = 20$, $n_x = 11$ and $n_\xi = 240$. Here, the conditions (C-I, C-II & C-III) are satisfied and therefore the solution is sparse in the stochastic domain. However, we do not have apriori a clear picture about the sparsity in the spatial domain for different random input dimensions (or within each PC coefficient). The reference values of $\mu(x)$ and $\sigma(x)$, needed for the evaluation of the measures of performance, are estimated by Monte Carlo integration (Robert and Casella, 2004) because they are not available in closed forms.

In Figs. 1b & 1c, we present the estimated marginal inclusion posterior probabilities of spatial and PC bases, in pairs, for [14D] and [40D] cases. Figs. 1e & 1f illustrate the significant scalar coefficients (of lower-order) for cases [14D] and [40D]. Only 73 out of 680 available PC bases in [14D] and 71 out of 850 available PC bases in [40D] are detected as significant by the proposed method according to the MPM rule. This shows that the proposed method based on the MPM rule has recovered sparse representations of the solutions in the stochastic domain, in both [14D] and [40D] cases. It is worth noting that, in the [40D] case, the PC

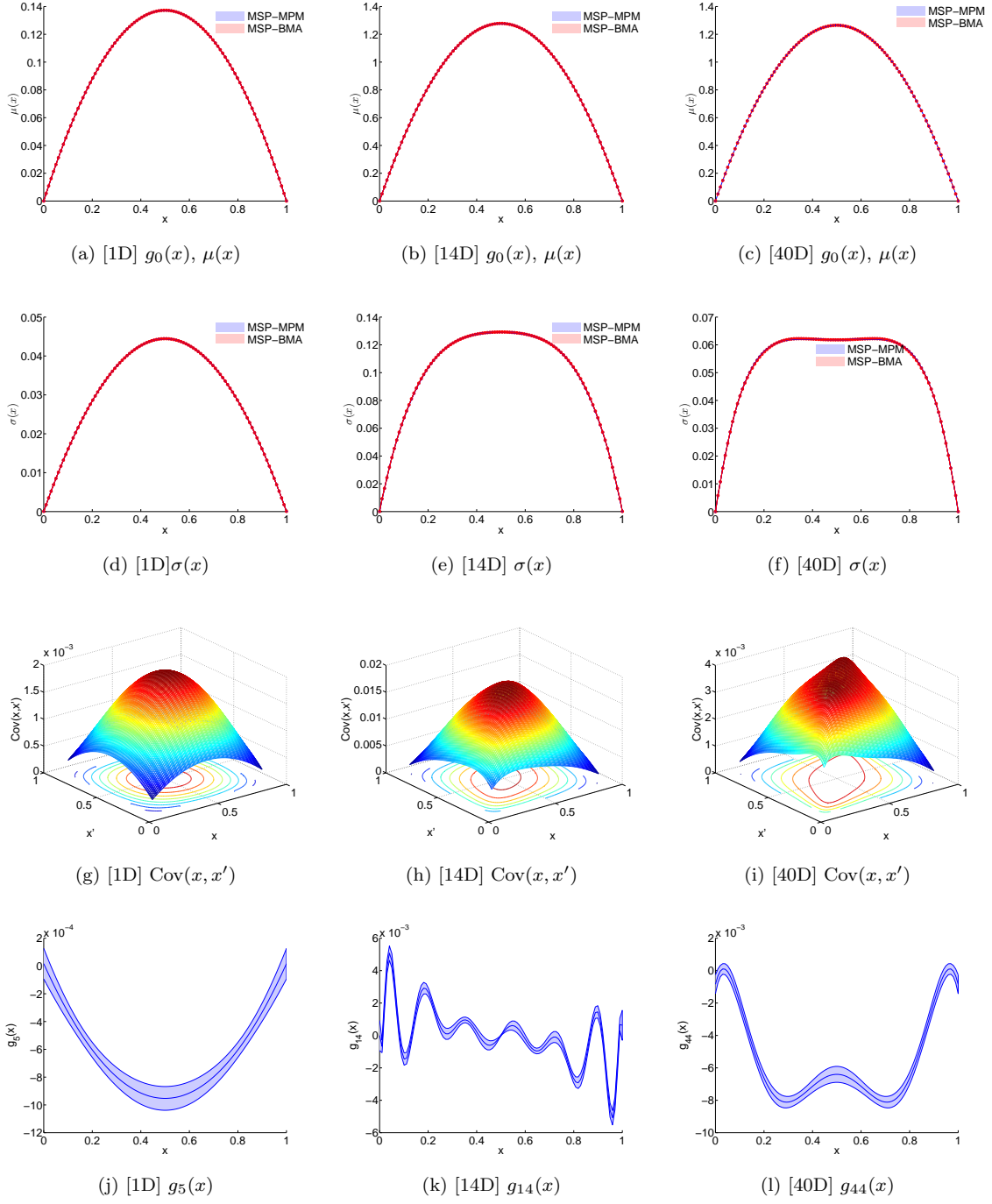


Figure 3: Estimates and 95%-CI of significant PC coefficients (panels a-c & j-l), the solution standard deviation (panels d-f) and covariance (panels g-i) as functions of the spatial input variable x . Notation: Mixed shrinkage (MSP), Bayesian model average (BMA) and median probability model (MPM).

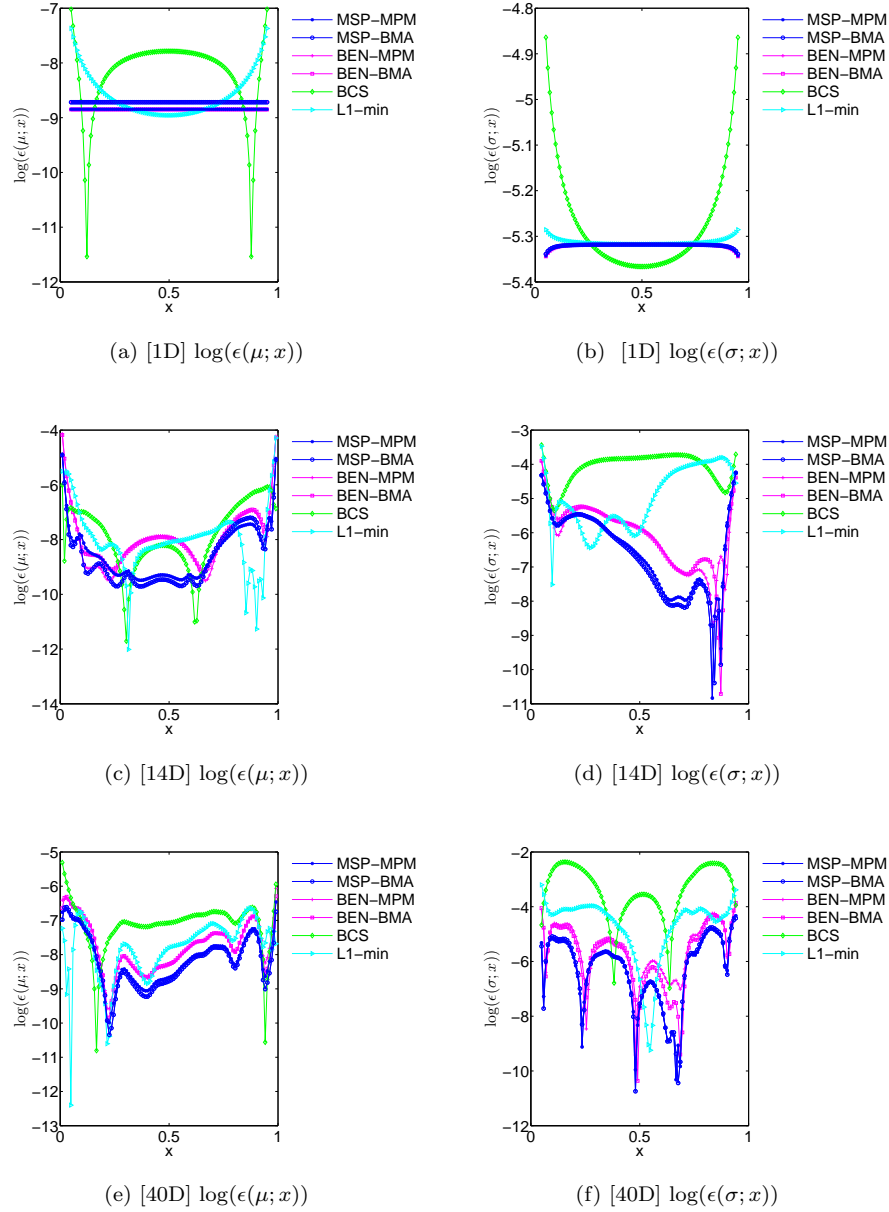


Figure 4: Relative errors of the solution mean $\epsilon(\mu; x)$ and standard deviation $\epsilon(\sigma; x)$, in the log scale, as functions of the spatial input variables. Approaches under comparison (by acronym): Mixed shrinkage (MSP), Bayesian elastic net (BEN), Bayesian compressive sensing (BCS) and l_1 -minimization (l_1 -min). Evaluation approaches for (MSP & BEN): Bayesian model average (BMA) and median probability model (MPM). Here, log denotes the natural logarithm.

bases associated to random dimensions between 23rd–40th were found to be statistically non-significant and so those random dimensions are statistically not important (Fig. 1f). The proposed gPC formulation in the Bayesian framework can provide detailed information regarding the sparsity of the solution in both stochastic and spatial domains. For instance, in Figs. 1e & 1f we observe that the sparsity in the spatial domain among

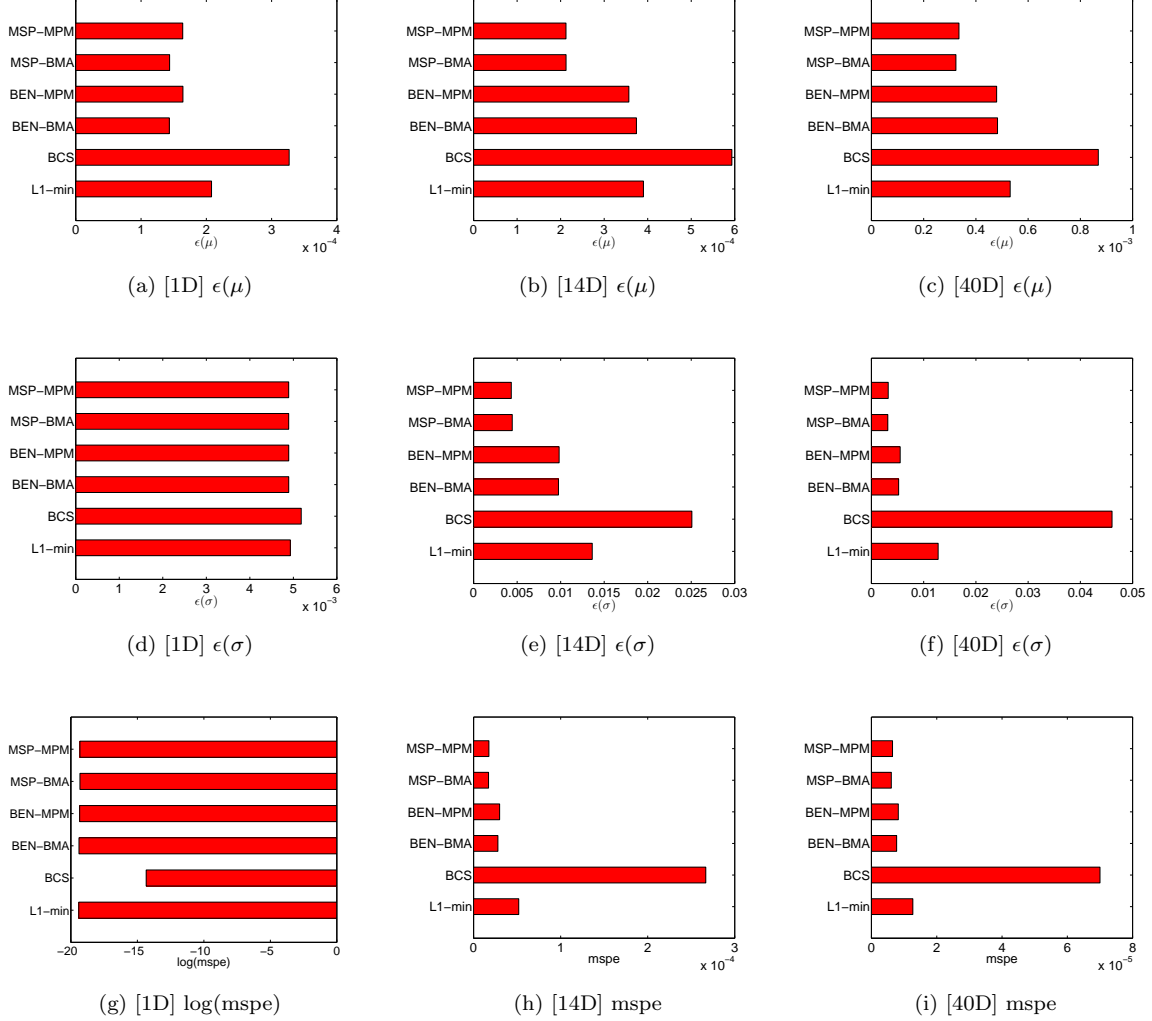


Figure 5: Comparative results concerning the total relative error of the solution mean $\epsilon(\mu)$, standard deviation $\epsilon(\sigma)$ and the mean squared predictive error mspe. Approaches under comparison (by acronym): The mixed shrinkage (MSP), Bayesian elastic net (BEN), Bayesian compressive sensing (BCS) and l_1 -minimization (l_1 -min). Evaluation approaches for (MSP, BEN): Bayesian model average (BMA) and median probability model (MPM). Here, log denotes the natural logarithm.

different PC coefficients (or random input dimensions) varies. In fact, higher-order PC coefficients appear to be more sparse than the lower-order ones. Moreover, we observe that higher-order PC coefficients (associated with random input interactions) tend to be adequately modeled by lower-degree polynomials. Given that PC coefficients reflect the effect of the associated PC bases (or random input dimensions) on the solution as functions of the spatial input, we observe that smaller-scale variation with respect to ξ , captured by higher-order PC bases, changes in a smooth manner over the spatial domain. This indicates that the solution is relatively smooth regarding x and ξ .

Interval estimation for the solution statistics and the PC coefficients can be performed. For representation purposes, the estimates and associated 95%-CI for the solution mean, standard deviation, covariance and PC coefficients are plotted over the spatial input domain in Figs. 3b, 3e, 3h & 3k for [14D] case and Figs. 3c, 3f, 3i & 3l for [40D] case. Figs. 3h, 3i, 3k & 3l refer to MPM evaluation as we have noticed no remarkable visual difference between MPM and BMA estimates. The plots highlight an advantage of the proposed method over existing compressive sensing approaches in gPC literature (Yang and Karniadakis, 2012; Doostan and Owhadi, 2011), which is that the PC coefficients and statistics of the gPC expansion can be recovered over the whole spatial domain at once, allowing results to be obtained at any spatial input value without the need to repeat the MC sampling and expansion evaluation each time. Moreover, the uncertainty regarding the solution statistics and PC coefficients can be quantified through their posteriors or credible intervals as shown in Figs. 3k & 3l.

The proposed method provides accurate enough results in high-dimensional cases. Meanwhile, the MSP procedure outperforms the other evaluating algorithms under comparison according to the performance measures considered as reported in Figs. 5b, 5e & 5h for the [14D] case and Figs. 5c, 5f & 5i for the [40D] case. MSP appears to present better performance than BEN due to its ability to directly address non-homogeneous group sparsity for different PC coefficients (or random input dimensions). We observe that MSP (both BMA and MPM) outperforms l_1 -min and BCS possibly because different PC coefficients present different sparsity in the spatial domain (as displayed in Figs 1e & 1f) and so different penalization shrinkage is required for each. Moreover, the probabilistic basis selection procedures (MSP, BEN) seem to outperform the standard compressive sensing (l_1 -min, BCS) algorithms. Thus, we observe empirically that the probabilistic basis selection approach provides a preferable mechanism for choosing significant basis functions compared to those of standard compressive sensing algorithms that mainly shrink all of the coefficients to zero. Although outperformed by MSP-BMA that accounts for uncertainty about bases significance, MSP-MPM results in mse and relative errors that are small in value and close to those of MSP-BMA in spite of the omission of

basis functions.

The relative errors for the solution mean and standard deviation, in log scale, are reported as functions of spatial input variable in Figs. 4c & 4d for the [14D] case and Figs. 4e & 4f for the [40D] case. We observe that MSP tends to present smaller relative errors in the biggest part of the spatial domain when compared to the other computational algorithms considered. Moreover, we observe that the lines of relative errors of MSP-MPM are closely above those of MSP-BMA over the whole spatial domain, with respect to the relative errors over the spatial domain. This indicates that MSP-MPM is a preferable alternative to MSP-BMA, in the gPC context, if the solution is sparse or the researcher needs to have a sparse representation of it without sacrificing in accuracy.

We study the performance of the proposed method, in [40D] cases: (i) for an increasing number of MC samples $n_\xi \in (200, 800)$ but a fixed number of spatial points $n_x = 11$ and (ii) for an increasing number of spatial points $n_x \in (10, 50)$ but a fixed number of MC samples $n_\xi = 300$ (Fig. 6). We compare the proposed methodology that involves modeling the PC coefficients as a series of Legendre bases and using (MSP, BEN, l_1 -min, BCS) against a 2-stage MC approach that involves evaluating the solution statistics at each spatial point individually and interpolating them over the spatial domain. Fig. 6a, 6b, 6d & 6e show that the proposed Bayesian method outperforms the 2-stage MC approach in terms of the total relative errors of the solution mean and standard deviation. When the gPC expansion is evaluated under the proposed gPC formulation (7), we observe that the accuracy of the expansion and its statistics improve as n_ξ and n_x increase with respect to the measures of performance considered (Fig. 6). This behavior is even more striking in Figs. 6b, 6e, 6c & 6f that show the decrease of $\epsilon(\sigma)$ and mspe as n_ξ and n_x increase. Given the proposed gPC formulation (7), Fig 6 shows that MSP produces better results for the evaluation of the expansion and its statistics than BEN, BCS and l_1 -min as n_ξ and n_x increase. Possibly, this is due to the MSP procedure's ability to consider heterogeneous sparsity in the spatial domain among different PC coefficients which are included in the estimator. In Figs. 6c & 6f, we observe that MSP outperforms the other algorithms under comparison in terms of mspe. Moreover, the mspe from MSP decreases quicker than those from the other algorithms under comparison as n_ξ and n_x increases.

5. Conclusions and future work

For the first time, in the gPC context, we proposed a new fully Bayesian method to efficiently obtain the spectral representation of a spatial random field, which can conduct spatial-stochastic basis selection and evaluation of gPC expansions when the number of the available basis functions is significantly larger

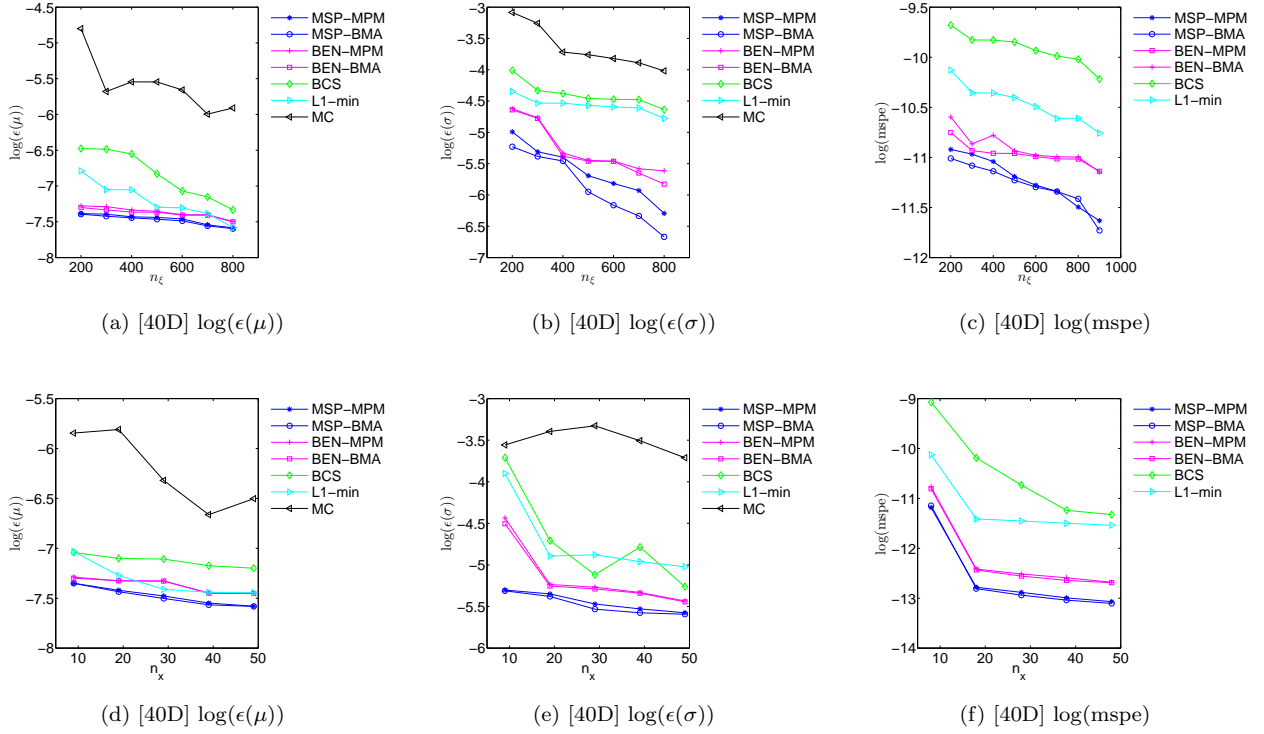


Figure 6: Total relative errors of the solution mean $\epsilon(\mu)$, standard deviation $\epsilon(\sigma)$ and total mean square predictive error mspe , in the log scale, for an increasing number of MC samples, $n_\xi \in (200, 800)$ and $n_x = 11$, (panels a & b) and increasing number of spatial points, $n_\xi = 300$ and $n_x \in (10, 50)$ (panels c & d). Approaches under comparison (by acronym): Monte Carlo (MC), mixed shrinkage (MSP), Bayesian elastic net (BEN), Bayesian compressive sensing (BCS) and l_1 -minimization (l_1 -min). Evaluation approaches for (MSP & BEN): Bayesian model average (BMA) and median probability model (MPM). Here, \log denotes the natural logarithm.

than the size of the training data-set. Hence, the gPC expansion is suitable for repeated use to obtain results economically at any new spatial input values without repeating the MC sampling or the algorithm to evaluate the expansion that existing compressive sensing methods in gPC require. We developed a fully Bayesian stochastic procedure, called mixed shrinkage prior (MSP) able to perform both basis selection and coefficient evaluation simultaneously. In general lines, the proposed methodology, we suggested, involves reformulating the traditional gPC expansion by modeling the PC coefficients as a series of basis functions and performing stochastic basis selection and evaluation of the coefficients by using our newly developed Bayesian MSP procedure. The method allows the recovery of possible sparse structures of the solution in stochastic and spatial domains. Due to its Bayesian nature, it is more informative than traditional l_1 -min approaches since it quantifies the uncertainty about significant bases and provides interval estimation via the associated posterior distributions. In addition, the MSP procedure developed can address heterogeneous sparsity in the spatial domain for different random input dimensions by compromising between LASSO and

Ridge regressions suitably.

The good performance of our method was exploited through applications to elliptic SPDEs with high-dimensional random diffusion coefficients, where the sparsity of the solution in the stochastic domain was guaranteed. We observed that the proposed method recovered a sparse representation of the solution in both stochastic and spatial domains adequately. Compared to a 2-stage MC method that performs evaluation of the solution statistics at several locations and interpolation of them over the spatial domain, the proposed Bayesian method presented more accurate results. Our empirical comparison showed that the proposed Bayesian MSP procedure is preferable to other existing algorithms under comparison, with respect to the measures of performance considered, especially in scenarios where PC coefficients present heterogeneous sparsity in the spatial domain for different random input dimensions.

Future work can focus on whether it is possible to reduce the number of available basis functions $\{\theta_b(\cdot)\psi_a(\cdot); (a, b) \in A \times B\}$ for the gPC expansion (7) by applying a truncation rule before performing the MSP procedure and without increasing the truncation error significantly. In fact, our examples imply that higher-order PC coefficients may be sparser in the spatial domain and adequately approximated by lower-degree polynomial series. The proposed method can be possibly extended to consider discontinuity by using binary tree partitioning (Chipman et al., 1998; Konomi et al., 2014) or capture smaller-scale variations, unexplained by the gPC part, by modeling the total truncation error term as a Gaussian process (O’Hagan, 1978; Bilonis et al., 2013). We believe that the former can lead to simpler gPC expansions while the latter can improve the estimates. This is ongoing work and the results will be presented in future publications.

Acknowledgments

This work was supported by the Applied Mathematics Program within the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR) as part of the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4). PNNL is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. We would like to thank Dr Kenneth Jarman (PNNL) for carefully proof-reading the manuscript. The research was performed using PNNL Institutional Computing, as well as the National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory.

References

Andrieu, C., Thoms, J., 2008. A tutorial on adaptive MCMC. *Stat. Comput.* 18 (4), 343–373.

- Babuška, I., Nobile, F., Tempone, R., May 2007. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* 45 (3), 1005–1034.
- Babuška, I., Tempone, R., Zouraris, G., 2004. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.* 42 (2), 800–825.
- Barbieri, M. M., Berger, J. O., 2004. Optimal predictive model selection. *Ann. Stat.* 32 (3), 870–897.
- Bayes, M., Price, M., 1763. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions* (1683-1775), 370–418.
- Bieri, M., Andreev, R., Schwab, C., 2009. Sparse tensor discretization of elliptic spdes. *SIAM Journal on Scientific Computing* 31 (6), 4281–4304.
- Bieri, M., Schwab, C., 2009. Sparse high order FEM for elliptic SPDEs. *Comput. Method Appl. Mech. Eng.* 198, 1149–1170.
- Bilionis, I., Zabaras, N., Konomi, B. A., Lin, G., 2013. Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification. *J. Comput. Phys.* 241, 212–239.
- Blatman, G., Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.* 230 (6), 2345–2367.
- Casella, G., Berger, R. L., 1990. *Statistical inference*. Vol. 70. Duxbury Press Belmont, CA.
- Cheney, E. W., 1966. *Introduction to approximation theory*. Vol. 3. McGraw-Hill, NY.
- Chipman, H. A., George, E. I., McCulloch, R. E., 1998. Bayesian CART model search. *J. Amer. Statist. Assoc.* 93 (443), 935–948.
- Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (1), 209–226.
- De Boor, C., 1978. *A practical guide to splines*. Vol. 27. Springer-Verlag New York.
- de Laplace, P., 1812. *Théorie analytique des probabilités*. V. Courcier.
- Deb, M. K., Babuška, I. M., Oden, J. T., 2001. Solution of stochastic partial differential equations using galerkin finite element techniques. *Comput. Method Appl. Mech. Eng.* 190 (48), 6359–6372.

- Doostan, A., Owhadi, H., Apr. 2011. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.* 230 (8), 3015–3034.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Ghanem, R. G., Spanos, P., 2003. *Stochastic finite elements: a spectral approach*. Dover publications, Mineola, NY.
- Gilks, W., Richardson, S., Spiegelhalter, D., 1996. *Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics. Chapman & Hall/CRC.
- Gilks, W. R., 1992. Derivative-free adaptive rejection sampling for Gibbs sampling. *Bayesian Statistics* 4.
- Gilks, W. R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41, 337–348.
- Gramacy, R. B., Lee, H., 2012. Cases for the nugget in modeling computer experiments. *Stat. Comput.* 22 (3), 713–722.
- Hans, C., Apr. 2010. Model uncertainty and variable selection in bayesian LASSO regression. *Stat. Comput.* 20 (2), 221–229.
- Hans, C., 2011. Elastic net regression modeling with the orthant normal prior. *J. Amer. Statist. Assoc.* 106 (496), 1383–1393.
- Hoeting, J. A., Madigan, D., Raftery, A. E., Volinsky, C. T., 1999. Bayesian model averaging: a tutorial. *Statistical science* 14, 382–401.
- Holmes, C., Mallick, B., 2001. Bayesian regression with multivariate linear splines. *J. R. Stat. Soc. Series B (Statistical Methodology)* 63 (1), 3–17.
- Jaynes, E. T., 2003. *Probability theory: the logic of science*. Cambridge university press.
- Jeffrey, H., 1961. *Theory of probability*. Clarendon Press, Oxford, 72.
- Ji, S., Dunson, D., Carin, L., 2009. Multitask compressive sensing. *IEEE T. Image Processing* 57 (1), 92–106.
- Ji, S., Xue, Y., Carin, L., 2008. Bayesian compressive sensing. *IEEE T. Signal Process.* 56 (6), 2346–2356.

- Karagiannis, G., Lin, G., 2014. Selection of polynomial chaos bases via bayesian model uncertainty methods with applications to sparse approximation of pdes with stochastic inputs. *Journal of Computational Physics* 259, 114–134.
- Konomi, B., Karagiannis, G., Sarkar, A., Sun, X., Lin, G., 2014. Bayesian treed multivariate gaussian process with adaptive design: Application to a carbon capture unit. *Technometrics* 56 (2), 145–158.
- Li, W., Lin, G., Zhang, D., 2014. An adaptive anova-based pckf for high-dimensional nonlinear inverse modeling. *Journal of Computational Physics* 258, 752–772.
- Lykou, A., Ntzoufras, I., Feb. 2012. On Bayesian LASSO variable selection and the specification of the shrinkage parameter. *Stat. Comput.*, 1–30.
- Ma, X., Zabaras, N., 2009. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.* 228 (8), 3084–3113.
- Ma, X., Zabaras, N., 2010. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *J. Comput. Phys.* 229 (10), 3884–3915.
- Madigan, D., Raftery, A., 1994. Model selection and accounting for model uncertainty in graphical models using occam’s window. *J. Amer. Statist. Assoc.* 89 (428), 1535–1546.
- Mathelin, L., Hussaini, M. Y., 2003. A stochastic collocation algorithm for uncertainty analysis. Tech. rep., NASA Contractor Report, Langley Research Center, Washington D.C.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087.
- Nobile, F., Tempone, R., Webster, C. G., 2008. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.* 46 (5), 2411–2442.
- O’Hagan, A., 1978. Curve fitting and optimal design for prediction (with discussion). *J. R. Stat. Soc. Series B* 40 (1), 1–42.
- O’Hagan, A., Forster, J., 2004. Kendall’s Advanced Theory of Statistics, volume 2B: Bayesian Inference, second edition. Vol. 2B. Arnold.
- Park, T., Casella, G., 2008. The Bayesian LASSO. *J. Amer. Statist. Assoc.* 103 (482), 681–686.

- Ripley, B. D., 1987. Stochastic simulation. John Wiley & Sons, Inc., NY.
- Robert, C. P., Casella, G., Jul. 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer.
- Roberts, G. O., Gelman, A., Gilks, W. R., 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.* 7, 110–120.
- Sargsyan, K., Safta, C., Najm, H., Debusschere, B., Ricciuto, D., Thornton, P., 2013. Dimensionality reduction for complex models via Bayesian compressive sensing. *International Journal for Uncertainty Quantification* 1 (1).
- Schwab, C., Todor, R. A., 2006. Karhunen - Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.* 217 (1), 100 – 122.
- Scott, J. G., Berger, J. O., 2010. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.* 38 (5), 2587–2619.
- Silverman, B. W., 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Stat. Soc. Series B (Methodological)*, 1–52.
- Stein, M. L., 1999. Interpolation of spatial data: some theory for kriging. Springer-Verlag, NY.
- Todd, J., 1963. Introduction to the constructive theory of functions. Birkhäuser Basel.
- Todor, R. A., Schwab, C., 2007. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.* 27 (2), 232–261.
- Wan, X., Karniadakis, G. E., 2005. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.* 209 (2), 617–642.
- Wan, X., Karniadakis, G. E., 2006. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.* 28 (3), 901–928.
- Wei, J., Lin, G., Jiang, L., Efendiev, Y., in press. Anova-based mixed msfem and applications in stochastic two-phase flows. *International Journal for Uncertainty Quantification*.
- Xiu, D., Jul. 2010. Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton University Press, Princeton, NJ.

- Xiu, D., Hesthaven, J. S., 2005. High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. 27 (3), 1118–1139.
- Xiu, D., Karniadakis, G. E., 2002. The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. 24 (2), 619–644.
- Yang, X., Choi, M., Lin, G., Karniadakis, G. E., 2012. Adaptive anova decomposition of stochastic incompressible and compressible flows. J. Comput. Phys. 231 (4), 1587–1614.
- Yang, X., Karniadakis, G. E., 2012. Reweighted l_1 minimization method for stochastic partial differential equations, reprint submitted to J. Comput. Phys.

Appendix

We give a rough sketch of the derivation of Eq. (12)-(19).

A summary of the Bayesian model is:

- The likelihood function is:

$$\mathcal{L}(u|c, \sigma^2) = \mathbf{N}_n(u|X^\top c, \mathbb{I}_n \sigma^2).$$

- The prior model is :

$$\begin{aligned} \Pr(\gamma_{a,b}|\rho) &= \text{Bernoulli}(\gamma_{a,b}|\rho), \forall(a,b); \\ \pi(c_{a,b}|\gamma_{a,b}, \sigma^2, \eta_a, \lambda_a) &= \begin{cases} \mathbf{N}(c_{a,b}|0, \sigma^2/\lambda_a) & , \text{ if } \gamma_{a,b} = 1, \eta_a = 1 \\ \text{DE}(c_{a,b}|0, \sigma/\lambda_a) & , \text{ if } \gamma_{a,b} = 1, \eta_a = 0, \forall(a,b); \\ \delta_0(c_{a,b}) & , \text{ if } \gamma_{a,b} = 0 \end{cases} \\ \pi(\sigma^2|a_\sigma, b_\sigma) &= \mathbf{G}(\sigma^2|a_\sigma, b_\sigma); \\ \pi(\eta_a|\varpi) &= \text{Bernoulli}(\eta_a|\varpi), \forall a; \\ \pi(\lambda_a|a_\lambda, b_\lambda) &= \mathbf{G}(\lambda_a|a_\lambda, b_\lambda), \forall a; \\ \pi(\rho|a_\rho, b_\rho) &= \text{Beta}(\rho|a_\rho, b_\rho); \\ \pi(\varpi|a_\varpi, b_\varpi) &= \text{Beta}(\varpi|a_\varpi, b_\varpi). \end{aligned}$$

The joint posterior distribution (computed according to the Bayes' theorem) density is such that

$$\begin{aligned}
\pi(\gamma, c, \eta, \rho, \varpi, \sigma^2, \lambda | u, X) &\propto \mathbf{N}_n(u | X^\top c, \mathbb{I}_n \sigma^2) \\
&\times \prod_{\forall (a,b)} \text{Bernoulli}(\gamma_{a,b} | \rho) \\
&\times \prod_{\forall (a,b)} (\gamma_{a,b} (\eta_a \mathbf{N}(c_{a,b} | 0, \sigma^2 / \lambda_a) + (1 - \eta_a) \text{DE}(c_{a,b} | 0, \sigma / \lambda_a)) + (1 - \gamma_{a,b}) \delta_0(c_{a,b})) \\
&\times \mathbf{G}(\sigma^2 | a_\sigma, b_\sigma) \\
&\times \prod_{\forall a} \text{Bernoulli}(\eta_a | \varpi) \times \prod_{\forall a} \mathbf{G}(\lambda_a | a_\lambda, b_\lambda) \\
&\times \text{Beta}(\rho | a_\rho, b_\rho) \times \text{Beta}(\varpi | a_\varpi, b_\varpi).
\end{aligned}$$

Regarding the derivation of Eq. (12)-(19). The full conditional posterior distribution density of $\pi(\gamma_{a,b}, c_{a,b} | u, X, \dots)$ is such that

$$\begin{aligned}
\pi(\gamma_{a,b}, c_{a,b} | u, X, \dots) &\propto \mathbf{N}(u | X^\top c, \mathbb{I}_m \sigma^2) \\
&\times (\gamma_{a,b} (\eta_a \mathbf{N}(c_{a,b} | 0, \sigma^2 / \lambda_a) + (1 - \eta_a) \text{DE}(c_{a,b} | 0, \sigma / \lambda_a)) + (1 - \gamma_{a,b}) \delta_0(c_{a,b})) \\
&\times \text{Bernoulli}(\gamma_{a,b} | \rho).
\end{aligned}$$

We consider $\pi(\gamma_{a,b}, c_{a,b} | u, X, \dots) = \Pr(\gamma_{a,b} | u, X, \dots) \times \pi(c_{a,b} | u, X, \gamma_{a,b})$. For the marginal $\Pr(\gamma_{a,b} | u, X, \eta_{a,\dots}) = \int \pi(\gamma_{a,b}, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}$, it is

$$\begin{aligned}
\Pr(\gamma_{a,b} = 1 | u, X, \eta_{a,\dots}) &= \frac{\int \pi(\gamma_{a,b} = 1, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}}{\int \pi(\gamma_{a,b} = 1, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b} + \int \pi(\gamma_{a,b} = 0, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}}; \\
&= \left[1 + \frac{\int \pi(\gamma_{a,b} = 0, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}}{\int \pi(\gamma_{a,b} = 1, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}} \right]^{-1}.
\end{aligned}$$

Note that $P_{a,b}^{(\gamma)} = \Pr(\gamma_{a,b} = 1 | u, X, \eta_{a,\dots})$. We distinguish cases for $\eta_a \in \{0, 1\}$, and find $\frac{\int \pi(\gamma_{a,b}=0, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}}{\int \pi(\gamma_{a,b}=1, c_{a,b} | u, X, \eta_{a,\dots}) dc_{a,b}} =$

$$\begin{cases} \left[1 + 2 \frac{1-\rho}{\rho} \frac{\sigma}{\lambda_a} / \left(\frac{\Phi(-\mu_{a,b}^- / s_{a,b})}{\mathbf{N}(0 | \mu_{a,b}^-, s_{a,b}^2)} + \frac{\Phi(\mu_{a,b}^+ / s_{a,b})}{\mathbf{N}(0 | \mu_{a,b}^+, s_{a,b}^2)} \right) \right]^{-1}, & \eta_a = 0 \\ \left[1 + \frac{1-\rho}{\rho} \sqrt{\frac{2\pi\sigma^2}{\lambda_a}} \mathbf{N}(0 | \mu_{a,b}^*, s_{a,b}^{2,*}) \right]^{-1}, & \eta_a = 1 \end{cases}.$$

For $\pi(c_{a,b} | u, X, \gamma_{a,b})$, we distinguish cases for $\eta_a \in \{0, 1\}$, and $\gamma_{a,b} \in \{0, 1\}$

$$\pi(c_{a,b} | u, X, \gamma_{a,b}) \propto \begin{cases} \mathbf{N}(u | X^\top c, \mathbb{I}_m \sigma^2) \times \text{DE}(c_{a,b} | 0, \sigma / \lambda_a) & , \eta_a = 0, \& \gamma_{a,b} = 1 \\ \mathbf{N}(u | X^\top c, \mathbb{I}_m \sigma^2) \times \mathbf{N}(c_{a,b} | 0, \sigma^2 / \lambda_a) & , \eta_a = 1, \& \gamma_{a,b} = 1, \\ \delta_0(c_{a,b}) & , \gamma_{a,b} = 0 \end{cases}$$

and find

$$\pi(c_{a,b}|u, X, \gamma \dots) \propto \begin{cases} (1 - w_{a,b})\mathbf{N}^-(c_{a,b}|\mu_{a,b}^-, s_{a,b}^2) + w_{a,b}\mathbf{N}^+(c_{a,b}|\mu_{a,b}^+, s_{a,b}^2) & , \eta_a = 0, \&\gamma_{a,b} = 1 \\ \mathbf{N}(c_{a,b}|\mu_{a,b}^*, s_{a,b}^{2,*}) & , \eta_a = 1, \&\gamma_{a,b} = 1 , \\ \delta_0(c_{a,b}) & , \gamma_{a,b} = 0 \end{cases}$$

where the calculations for (13), (14), (15), (16), (17), and (12) are straightforward, and the same as those of the standard Bayesian Ridge and LASSO regressions using slap-and-spike priors (Gilks et al., 1996; Hans, 2010). This is because the case $(\eta_a = 1, \&\gamma_{a,b} = 1)$ leads to the standard Bayesian Ridge regression using slap-and-spike priors, while $(\eta_a = 0, \&\gamma_{a,b} = 1)$ leads to the standard LASSO regressions using slap-and-spike priors.